Reichen, C., Hansen, S., Forzani, C., Honegger, A., Fleishman, S. J., Zhou, T., Parmeggiani, F., Ernst, P., Madhurantakam, C., Ewald, C., Mittl, P. R. E., Zerbe, O., Baker, D., Caflisch, A., & Plückthun, A. (2016). Computationally Designed Armadillo Repeat Proteins for Modular Peptide Recognition. *Journal of Molecular Biology*, *428*(22), 4467-4489. https://doi.org/10.1016/j.jmb.2016.09.012

Title: **Computationally Designed Armadillo Repeat Proteins for Modular Peptide Recognition.**

Authors: Christian Reichen[a], Simon Hansen[a], Cristina Forzani[a], Annemarie Honegger[a], Sarel J. Fleishman[b,c], Ting Zhou[a], Fabio Parmeggiani[a,e], Patrick Ernst[a], Chaithanya Madhurantakam[a,f], Christina Ewald[d], Peer R. E. Mittl[a], Oliver Zerbe[d], David Baker[b], Amedeo Caflisch[a] and Andreas Plückthun[a,*]

Affiliations:

[a] Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland

[b] Department of Biochemistry, University of Washington, Seattle, USA

[c] Current address: Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

[d] Department of Chemistry, University of Zurich, 8057 Zurich, Switzerland

[e] Current address: Life Sciences Building, University of Bristol, Bristol BS8 1TH, United Kingdom

[f] Current Address: Department of Biotechnology, TERI University, 10, Institutional Area, New Delhi 110070, India

* Corresponding author.

Address: Department of Biochemistry,

University of Zurich, Winterthurerstrasse 190,

8057 Zurich, Switzerland.

Phone: +41 44 635 55 70.

Fax: +41 44 635 57 12

E-mail: plueckthun@bioc.uzh.ch

Document: pages 39; figures 10; tables 3

Supplementary: *CReichen_Rosetta_SupplementaryText.docx*

supplementary pages 11; supplementary figures 9; supplementary tables 4

Coordinates: Coordinates and structure factors have been deposited at the PDB under the following entry codes:

4D4E ($Y_{III}(Dq)_4C_{PAF}$), 4D49 ($Y_{III}(Dq.V1)_4C_{PAF}$)

Abstract

Armadillo repeat proteins (ArmRP) recognize their target peptide in extended conformation and bind, in a first approximation, two residues per repeat. They may thus form the basis for building a modular system, in which each repeat is complementary to a piece of the target peptide. Accordingly, preselected repeats could be assembled into specific binding proteins on demand and thereby avoid the traditional generation of every new binding molecule by an independent selection from a library. Stacked armadillo repeats, each consisting of 42 amino acids arranged in three α-helices, build an elongated superhelical structure. Here, we analyzed curvature variations in natural ArmRPs, and identified a repeat pair from yeast importin-α as having the optimal curvature geometry to be complementary to a peptide over its whole length. We employed a symmetric *in silico* design to obtain a uniform sequence for a stackable repeat while maintaining the desired curvature geometry. Computationally designed armadillo repeat proteins (dArmRPs) had to be stabilized by mutations to remove regions of higher flexibility, which were identified by molecular dynamics (MD) simulations in explicit solvent. Using an N-capping repeat from the consensus-design approach, two different crystal structures of dArmRP were determined. Although the experimental structures of dArmRP deviated from the designed curvature, the insertion of the most conserved binding pockets of natural ArmRPs onto the surface of dArmRPs resulted in binders against the expected peptide with low nanomolar affinities, similar to the binders from the consensus-design series.

Keywords

Armadillo repeat protein, Computational protein design, Rosetta, Symmetric design, Molecular dynamics, Peptide binding, X-ray crystallography

Abbreviations used

ANS, 1-anilino-8-naphthalene-sulfonate; AU, asymmetric unit; dArmRP, computationally designed armadillo repeat protein; cArmRP, consensus armadillo repeat protein; nArmRP, natural armadillo repeat protein; CD, circular dichroism; ELISA, enzyme linked immunosorbent assay; IMAC, immobilized metal-ion affinity chromatography; MALS, multi-angle light scattering; MD, molecular dynamics; MRE, mean residue ellipticity; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SEC, size-exclusion chromatography.

Introduction

Specific protein recognition is essential for many physiological processes, and forms the basis of a number of procedures routinely used in research, diagnostics and therapeutics. Still, the generation of binding reagents is time-consuming and has to be carried out for each desired target individually, either for monoclonal antibodies by immunization or for recombinant antibodies and alternative binding scaffolds by selection from a suitable library. A modular recognition of targets would allow using the same binding surface in multiple contexts and speed up research and development by reducing design and selection steps.

The complex features of globular protein surfaces usually prevent modular binding, but peptides are ideal targets for specific recognition by defined units, as in their extended form they constitute regularly spaced structural features. As "peptides" we refer not only to short stretches of amino acids but also to unstructured regions of proteins, such as termini, loops or linkers between domains. Peptide-protein interactions are found in many highly dynamic cellular networks involved in signaling, regulation and protein trafficking [1, 2] and represent about 15-40% of all interactions in the cell [3]. For many applications involving recognition of such proteins or epitopes on denatured or digested forms of folded proteins, a general peptide-binding scaffold will be particularly valuable, if it can provide a modular and specific recognition of the peptide primary sequence.

Among peptide-binding scaffolds, armadillo repeat proteins (ArmRP) were found to form a stable framework with 4-12 repeats and provide a constant binding mode for peptides in extended conformation, suitable for the generation of specific modular peptide binders (reviewed in [4]). Both natural (nArmRPs) and designed ArmRPs (dArmRPs) consist of several internal repeats, stacked in tandem on each other to form an elongated α-solenoid protein with a continuous hydrophobic core, having specialized capping repeats at the N- and C-termini. Each armadillo repeat (ArmR) unit, composed of 42 amino acids and folded into three α-helices (H1, H2, H3) in a spiral staircase mode, binds approximately two consecutive amino acids of the peptide. This is achieved by a conserved asparagine residue, N^{37} (superscripted numbers refer to the positions within the repeat) on each repeat, which fixes the peptide backbone by binding to every second peptide bond through bidentate hydrogen bonds, and by several surface residues forming pockets that interact with the peptide side-chains [5].

In both major subfamilies of nArmRPs, importin-α and β-catenin, the peptides are bound in an antiparallel orientation orientation (N- to C-terminal directions of protein and peptide run in opposite directions). The binding groove is built by a parallel arrangement of helix H3 of each repeat. In nArmRPs the conserved binding mode is limited to three consecutive repeats (Figure 1A) since the curvature is not constant across the entire protein. Thereby, the peptide units fall out of register with the armadillo repeats. In importin-α, two separated negatively charged binding sites (major and minor binding site) are formed by repeats B-D and F-H, respectively, and can bind a bipartite nuclear localization sequences (NLS), with the typical sequence $KRx_{10-12}K+x+$ ("+" denoting any positively charged residue [6]). Therein the two positively charged residue clusters are separated by a linker of 10-12 amino acids. In contrast to importin-α, β-catenin has only one binding site, which is positively charged, and the conserved binding is restricted to an area between repeats E-I.

This limited conserved binding of consecutive repeats in nArmRPs can be well explained by different curvatures found between neighboring repeat pairs. The analysis of nArmRP has shown that the sequence similarity between repeat units reaches only about 30%, and most differences — in length and residue composition — are found in the loops connecting the more conserved α-helices [7]. This low similarity is reflected in structural differences between repeats and accordingly in curvature variations between repeat pairs. In order to obtain a modular peptide-binding scaffold curvature should be uniform over the whole protein and fit to the register given by the unit length of the peptide backbone.

A consensus-based design approach, which had been applied for other repeat proteins [8] and also for ArmRPs [5, 9, 10] (named consensus ArmRPs or cArmRPs), is expected to yield a uniform curvature, although it may not necessarily match the exact geometry desired for binding the dipeptide units. Therefore, an *in silico* design approach was developed in this work, based on a geometrically optimal curvature template. Using this template, the relative orientations between subsequent repeats were extracted and imposed as symmetric modeling constraints during backbone and side chain sampling simulations using the Rosetta software suite [11]. Similar design protocols have been used for the computational design of repeat proteins, first with sequence and structural information obtained from natural repeat protein families [12, 13] and then for *de novo* designed repeat proteins with open [14] and closed [15] architectures. Using such approaches, typically >50% of the designed constructs

can be expressed as soluble, folded, monomeric proteins and determined structures agree well with the design models (typical RMSD of Cα atoms 0.5 – 2.5 Å).

Next, protein regions of higher flexibility were assessed by molecular dynamics (MD) simulations and more stable variants were engineered. Additional N-cap engineering allowed us to obtain an X-ray structure of a computationally designed ArmRP (dArmRP), which resembled the original template model with an RMSD of about 1.9 Å. However, we found a deviation from the intended curvature. Nonetheless, surface modifications that had been grafted from nArmRPs-peptide complexes enabled the generation of binders against positively charged peptides with low nM affinities.

The structure of the dArmRP - $(RR)_5$ peptide complex revealed that the peptide is bound in an antiparallel orientation along the dArmRP binding surface, as intended. This result demonstrates that, although the exact desired curvature was not achieved, binding in a modular manner was still achieved for three consecutive repeats in the case of the dArmRP scaffold.

Results

Superhelical curvature of natural Armadillo Repeat proteins

For the development of a modular ArmRP, the superhelical curvature of the protein must match the distances found in the peptide in its bound conformation across many peptide units, that is, the location of each protein repeat must be exactly in register with the peptide bonds. To analyze this correlation between the peptide and repeat geometry, we described the superhelical fold of ArmRPs using helical symmetry parameters. We characterized the radius (r), rise (h) and angle (2·Ω), which together describe the positions of internal repeats (at its center of mass (CoM)) on a helix around a central axis (Supplementary Figure S1). In the conserved binding mechanism, exemplified by the major binding site of yeast importin-α [16] (Figure 1A), the position of every second peptide bond is defined through the interaction of a double hydrogen bond with N^{37} in each Arm repeat. Accordingly, a specific ArmRP geometry provides a defined distance between a Cα-atom of an amino acid (P) of the bound peptide and the Cα-atom two amino acids C-terminal to it (P+2)).

When the peptide is bound in an extended conformation, the Cα(P/P+2)-distance should be 6.7-7.0 Å (Figure 1A). This Cα(P/P+2)-distance is observed for bound peptides of importin-α, which assume an

extended conformation within the major and minor binding site. This distance was also predicted by calculations and modeling of a peptide in relaxed β-strand conformation, including favored rotation angles and bond lengths. Note that the distance between two neighboring Cα atoms is constant at 3.8 Å because of the rigidity of the peptide bond, and, because of the tetrahedral angle linking to units, the Cα(P/P+2)-distance could maximally reach 7.0 Å.

In total, 36 peptide-bound structures of importin-α, which has overall 10 repeats, were analyzed geometrically by applying helical symmetry parameterization using Rosetta (Supplementary Figure S1). Repeat pairs containing N- or C-terminal capping repeats were excluded from the analysis, because of larger variations in sequence length and composition, compared to internal repeats. It was found that only repeat pair CD (from mouse and human importin-α) and GH (from yeast) display a curvature that places the Cα(P/P+2)-distance in the optimal range (Figure 1B). All other repeat pairs possess curvatures resulting in a larger Cα(P/P+2)-distance that would be inconsistent with modular peptide binding across a large number of repeats. We note that some natural ArmRPs solve this curvature inconsistency by recognizing several stretches of the peptide that are separated in sequence, such that this part acts like a linker, as described above for the bipartite NLS binding to importin-α.

Ideal Armadillo Repeat protein curvature for modular peptide binding

For the design of ArmRP possessing Cα(P/P+2)-distances suitable for peptide binding the GH repeat geometry was chosen over the CD repeats since GH repeats have lower Cα(P/P+2)-distances and were considered to be more generic, as short distances were observed in all structures. Finally, we focused on the GH repeat geometry originating from yeast importin-α with optimal Cα(P/P+2)-distances between 6.5-6.8 Å. To obtain a uniform multi-repeat curvature template with 12 repeats according to the GH repeat pair geometry (and to enforce a small Cα(P/P+2)-distance), copies of the GH repeat pair (PDB ID 1EE4 [17]) were superimposed repeat-wise (Supplementary Figure S2). The GH backbone model provided an idealized ArmRP curvature compatible with binding a peptide over the whole length. The optimal curvature is characterized by a large radius (r = 15.7 Å), a small rise (h = 6.2 Å) and an intermediate angle (2·Ω = 29.3°) in comparison to other curvatures found in yeast importin-α (Figure 2A and Supplementary Figure S2D). With these curvature parameters, a bound peptide is expected to have a Cα(P/P+2)-distance of 6.5 Å.

In silico design

In silico design was used to find an optimal amino acid sequence for the dArmR that would support the desired curvature. Internal repeats were constrained to adopt the same primary sequence, side-chain, and backbone conformations by using symmetric sequence design and conformation sampling during all modeling moves. Symmetric structure prediction [18] and design has been used extensively in Rosetta [11] yielding atomic-accuracy predictions for large homomeric oligomers, designed cage-like assemblies or repeat proteins [12-15, 19-21]. Our calculations were restricted to three internal repeats and two capping repeats. Since the modeled subunits are identical, the repeat protein could be extended indefinitely. We additionally enforced the N^{37} residues, because they are crucial for binding the peptide in each repeat. All other residues in the ArmRP were allowed to be mutated (to all amino acids except cysteine) using the Rosetta all-atom energy function (score12), which is dominated by Lennard-Jones, hydrogen-bonding, and implicit-solvent interactions [18].

Capping repeats were based on the computationally designed internal repeat. This was achieved by exchanging exposed hydrophobic residues in the capping repeats to hydrophilic ones (N-cap: $A^{12}E$, $P^{15}Q$, $L^{19}W$, $V^{27}T$ and $A^{34}Q$; C-cap: $V^{8}E$, $V^{17}E$, $L^{20}Q$, $A^{32}Q$, $A^{36}E$ and $A^{39}N$), as had been done for designed ankyrin repeat proteins [22] and designed ArmRPs from the consensus-design series [10].

Several positions in internal and capping repeats did not converge to a unique identity. Frequency in multiple sequence alignments and helical propensity were used to select residues for these positions. In the Rosetta models, positions 14, 15, 17 and 32 in the internal repeat (D-type), positions 17 and 32 for the N-cap and positions 14, 15 and 37 for the C-cap allowed for several alternative residues (Figure 3 and Supplementary Figure S3). In the N-cap, Trp^{19} was introduced for practical reasons, to determine the protein concentration by UV absorbance at 280 nm. F^{38} was preserved in the C-cap as observed in natural armadillo repeat proteins, while the neighboring residues were redesigned using Rosetta.

The Rosetta models were analyzed by molecular dynamics (MD) simulations, which revealed instability of helix H3, probably due to the introduction of Ser residues with unfavorable helical propensity at positions 33 and 36. These positions were replaced by Ala residues, and an increase of stability was observed in MD simulations.

Experimental validation of computationally designed ArmRPs

We systematically tested 9 combinations of *in silico* designed variants experimentally (Supplementary Table ST1). Each construct consisted of an N-terminal capping repeat, four internal repeats and a C-terminal cap. ORFs were initially assembled as reported previously [5], but a faster one-day-multi-fragment ligation assembly protocol could be developed (Supplementary Figure S4). Proteins were purified by a single step of immobilized metal-ion affinity chromatography (IMAC) with yields of up to 80 mg per 1 L of *E. coli* culture. Purified proteins were tested for their biophysical properties, i.e. monomeric behavior, amount of secondary structure, accessibility of the hydrophobic core and chemical stability.

Construct $N_V(D_{SPVA})_4C_{PAF}$ was identified as monomeric and folded, with — compared to other variants — lower ANS binding, higher melting temperature and, importantly, more cooperative unfolding, as measured by GdnHCl-induced unfolding. This construct was therefore chosen as the basis for further engineering (Figure 4 and Supplementary Figure S3, S5, S6). This variant contains N- and C-terminal capping repeats of type N_V and C_{PAF}, respectively, and four internal repeats of type D_{SPVA} (subscripts describe the variable residues that were experimentally found to be superior (see below), cf. Figure 3 for sequence positions). For simplicity, $N_V(D_{SPVA})_4C_{PAF}$ will be named CAR0 (computationally designed ArmRP version 0). Additionally, CAR0 was also identified by 1D ^{1}H-NMR experiments as a promising candidate, and 2D-[^{15}N,^{1}H]-HSQC spectra confirmed that CAR0 is stable and structured (Supplementary Figure S7).

Stabilization of dArmRP for structure determination

Stabilization by MD simulations. Several attempts to obtain crystals of construct $N_V(D_{SPVA})_xC_{PAF}$ with 1-10 internal repeats failed, although all proteins had favorable biophysical properties (Supplementary Figure S8). Therefore, the Rosetta model of CAR0 was used as starting structure for multiple explicit solvent MD runs (of 0.5 to 2 μs each) to identify regions with high flexibility and stabilize them, as described before [9]. The MD trajectories showed that the overall fold of model CAR0 was preserved, with an average root mean square fluctuation (RMSF) of 0.4 Å for the C_α carbon atoms (the RMSD is given in Supplementary Table ST2). Higher conformational instability was observed for the caps with average RMSF values between 0.9-1.0 Å. The plasticity of the loop-residues was higher than the helix-residues with RMSF values of 0.96 Å and 0.74 Å, respectively.

On the basis of the RMSF values calculated along the MD sampling of CAR0, six mutations were introduced in each internal repeat to reduce fluctuations, resulting in the internal repeat Dq (Figure 3). An additional set of independent simulations (of 0.5 to 2 μs each) were started from the mutated protein (with Dq internal repeats) and showed lower RMSF profiles than CAR0. Guided by the MD simulation results, the full-length construct CAR1 ($N_V(Dq)_4C_{PAF}$) was produced, and yielded a monomeric and well-folded protein (Figure 4). In comparison to CAR0, the stability increased by 0.6 M GdnHCl (Table 1). However, the increase in stability during chemical unfolding did not correlate with thermal unfolding data. We attribute this fact to an unusually high thermal (but not chemical) stability of the internal repeats D_{SPVA}. To guide the stability improvement of the dArmRP constructs we relied on chemical denaturation data since this procedure, in contrast to thermal denaturation, was able to fully unfold the proteins (as judged by CD spectroscopy) and showed a clear, cooperative transition. Crystallization attempts of CAR1 were, however, without success.

Replacement of the N-cap for crystallization. Crystal structures of designed ArmRPs were so far only determined for consensus-based ArmRPs with N-caps rationally designed (named Y_{II} and Y_{III}) [10]. In order to analyze the effect of the Y_{III}-type capping repeat on dArmRPs in terms of biophysical properties and for their ability to form crystals, protein CAR2 ($Y_{III}(Dq)_4C_{PAF}$) was produced. The introduction of the Y_{III} cap increased the stability of CAR2 by 0.2 M GdnHCl, while thermal stability remained nearly identical (Figure 4 and Table1). Although the level of ANS binding increased compared to CAR1, it is similar to the well-folded consensus-based ArmRP $Y_{III}M_4A_{II}$ [10]. Overall, protein stability could be increased from CAR0 to CAR2 by a shift of 0.8 M GdnHCl in denaturation midpoint, which resembles the most stable dArmRP so far.

Structure of computationally designed ArmRP CAR2

CAR2 crystallized at pH 5.5 in 0.1 M sodium acetate, 0.3 M sodium cacodylate and 25% PEG 2K MME and diffracted to 2.0 Å resolution (Table 2). The structure was determined by molecular replacement. The two molecules in the asymmetric unit of CAR2 are aligned front-to-front by their C-terminal concave binding sites, burying a large surface area of 1682 $Å^2$ (Figure 5). The right-handed superhelical structure of each molecule has an overall dimension of 60 x 30 x 20 Å. The lowest temperature factors were observed for the internal repeats ($<B_{N\text{-}cap}>$ = 33.48 $Å^2$, $<B_{Internal}>$ = 27.09 $Å^2$ and $<B_{C\text{-}cap}>$ = 42.30 $Å^2$), as previously observed for consensus-design ArmRPs [10, 23] and other α-solenoid proteins [24-26] (Table 3). The largest temperature factors were found in the C-terminal

capping repeat, especially within the loop connecting helices 2 and 3, that also displays no crystal contacts.

Structural comparison of dArmRPs. Both molecules from the asymmetric unit showed rather high structural variations (Figure 5B). A comparison of chains A and B of CAR2 resulted in a RMSD of 1.1 Å (Cα residues from the N-cap (L13) to the C-cap (A247)). The comparison of each individual internal repeat to the corresponding one in the chain B from the asymmetric unit revealed that they superimpose with a RMSD of 0.5 Å. The major exception is the loop connecting helices 2 and 3 in the internal repeat 3 of chain B. However, these different loop conformations do not explain the large RMSD for the superposition of the whole chains, which stems from small curvature variations between the repeat pairs.

Comparison of the experimental CAR2 structure with the designed model. The structure of CAR2 resembled the originally designed Rosetta-based model with RMSD of 1.8 Å ± 0.2 Å (Cα of four internal repeats, averaged over both molecules of the asymmetric unit). Although this value is usually acceptable for a general design approach, the analysis of the curvatures indicated a significant deviation from the initial design (Figure 2). The model CAR0, which was obtained by Rosetta repacking based on the curvature of the GH repeat from importin-α (PDB ID: 1EE4), shows parameters (rise, radius and angle) that result in a small Cα(P/P+2)-distance of a modeled peptide. Accordingly, the overall shape can be described as a short and wide cylinder with a medium-sized angle between neighboring repeats (Figure 2A). The structure of CAR2 deviates from the model-curvature and can be described as a tall (large rise) and thin (small radius) cylinder with a large angle between neighboring repeats (Figure 2B). Accordingly, we found that the expected Cα(P/P+2)-distances of 8.2 ± 0.8 Å were significantly larger in CAR2 than in the initial design (Cα(P/P+2)-distance of CAR0: 6.4 Å). This difference between Cα-distances of 1.8 Å indicates that, although the design approach resulted in stable and typical α-solenoid folded proteins, the curvature of the apo-CAR2 structure is not meeting the requirements for a perfectly modular peptide-binding scaffold.

In contrast to natural ArmRPs, the curvature characteristics of repeat pairs from dArmRPs (Figure 2B and C) are more uniform, which is expected due to their identical sequence in each internal repeat. However, due to the observed structural difference between the molecules within the asymmetric unit, the uncertainties in the parameters in dArmRPs, e.g. the Cα(P/P+2)-distances with standard deviations of 0.8 Å, is still rather large.

Surface redesign for Peptide Binding

The strongest affinities of natural ArmRPs were reported to be in the range of 20 nM for importin-α binding to NLS peptides [27, 28]. From an initial version of a consensus ArmRP library, a binder to neurotensin was selected with a K_d of 7 μM [29], but the binding mode was different from the intended canonical binding [30]. More recently, picomolar affinities for the interaction between the designed ArmRP $Y_{III}M_6A_{II}$ and the $(KR)_4$-peptide have been measured [31]. In contrast, no significant peptide binding was detected for CAR2. This is not surprising, as surface residues, potentially involved in binding, were neither selected nor engineered for binding yet.

Therefore, a design of the dArmRP binding site was undertaken, inspired by crystal structures of ArmRPs in complex with the target peptide. Similar approaches have been used to graft binding pockets onto the scaffold of TPR domains [32-34]. The analysis of 36 structures of nArmRPs binding to NLS peptides revealed two highly conserved binding pockets at the major (P2 and P3) or the minor binding site (P1' and P2') [4], both binding the side chain of positively charged amino acids (Lys and Arg). Although the major and minor binding sites are similar, we focused on the minor binding site. In nArmRPs, binding pocket P1' is formed mainly by residues D^1, T^4 and A^{40}, while P2' is formed by E^{30} W^{33} and T^{40*} (the asterisk indicates a position in the preceding repeat). Although position 40 is located distantly from the center of the binding pocket P2', T^{40*} in P2' is highly conserved and forms one hydrogen bond to the backbone oxygen of the bound peptide. In order to allow occupation of both pockets over several repeats, we decided to place Ser at position 40, which can be seen as a compromise between Ala and Thr, to accommodate the requirements for P1' and P2'. Surface mutations were introduced in newly produced proteins CAR2.V1 – CAR2.V4 (sequences are given in Figure 3) to mimic pockets P2' alone, or P1' in combination with P2'. Specific binding to positively charged peptides (e.g. $(KR)_4$ and $(KR)_5$) was qualitatively detected by ELISA for all surface variants, whereas only background binding was observed for protein CAR2. High background binding to the (KR)-peptides is expected due to cross-reactivity of the detection antibodies as illustrated by the no protein control (Figure 6).

In order to quantify protein affinity, fluorescence anisotropy assays were performed. Affinities were determined towards four different positively charged peptides ($(RR)_5$, $(RR)_4$, $(KR)_5$ and $(KR)_4$), fused to sfGFP as fluorescence marker. These peptides were chosen, because during pocket design no absolute preference for one of them could be deduced: within the analyzed 36 structures, pockets P1'

: P2 were found to be occupied by K/R with a frequency of 82%/6% : 97%/3%, while the frequencies for P2' : P3 were 9%/91% : 44%/44%, respectively. The His-tag of all surface-engineered dArmRPs was proteolytically removed to ensure an accessible binding groove (referred to as e.g. CAR2.V1_nohis). The affinity of the parental construct CAR2_nohis could only be approximately determined, since its low affinity to all peptides (>10 µM) would require very high protein concentrations that might interfere with assay conditions.

All the surface-engineered proteins gave rise to affinities in the nanomolar range, depending on the peptides tested (Table 1). The tightest interaction was observed for CAR.V2_nohis towards $(RR)_5$ (K_d 2.2 ± 0.1 nM), which corresponds to an increase in affinity of at least 4500-fold compared to CAR2_nohis. All surface-engineered binders show an increased affinity if the peptide is prolonged ($(RR)_4$ vs. $(RR)_5$ or $(KR)_4$ vs. $(KR)_5$), as previously determined for consensus ArmRP [31]. The effect is more pronounced for the $(KR)_n$ peptides, where one additional (KR)-unit leads to an 11 to 16-fold affinity increase for different proteins, while one (RR)-unit increases the affinity 4-6 fold. The additional mutation $K^{29}Q$, introduced in CAR.V1_nohis yielding CAR.V2_nohis, was designed to prevent the charge neutralization of E^{30} (Figure 7C), which might therefore contribute to side chain binding. This change resulted in an affinity increase by a factor of ca. 10 towards all peptides. The introduction of the second binding pocket P1' (mutation $N^{1}D$ and $I^{4}T$) only sometimes increased the affinity, e.g. CAR2.V3_nohis compared to the parental CAR2.V1_nohis binding to $(KR)_n$ peptides. However, in other instances the affinity did not change much (e.g. for CAR2.V3_nohis binding to $(RR)_n$ peptides) or was even decreased (e.g. for CAR2.V4_nohis compared to CAR2.V2_nohis when binding to all peptides). Therefore, contributions of individual mutations are not generally additive, and this point will require further investigations.

All constructs CAR2.V1-V4_nohis were monomeric, as measured by MALS, and eluted, with the exception of CAR2.V4_nohis, as a single symmetric peak, similar to CAR2_nohis (Figure 6). Although the modification of the surfaces did not alter the secondary structure, as characterized by the mean residue ellipticity (MRE) in CD spectra, it resulted in a decrease of stability by 0.3 – 1.4 M GdnHCl or by 10-22°C (change of denaturation midpoints) (Table 1).

Structure of CAR2.V1 complexed with peptide $(RR)_5$

For the structural investigation of the peptide-protein-interactions, surface variants were set up for crystallization with peptides $(KR)_5$ or $(RR)_5$. A structure of CAR2.V1 in complex with an $(RR)_5$ peptide was determined at 2.1 Å resolution by molecular replacement using the previously determined crystal structure of CAR2.

During the refinement, a poly-arginine peptide was modeled into the density observed at the binding site proving that each protein is binding one peptide. From the bound peptides, only nine amino acids were resolved (Arg1-9 in chain A and C, Arg2-10 in chain B and D). Additional density could be filled with single Arg residues, indicating that alternative binding conformations are possible. Always two molecules of the structure CAR2.V1 are aligned front to front, and with a parallel orientation to each other (Figure 8A). With this front-to-front orientation of two ArmRPs, N- and C-terminus of two antiparallel-bound peptides bound along the inner binding site are positioned closely to one another. Analogously to the CAR2 structure, the lowest temperature factors for CAR2.V1 were observed for the internal repeats, while the highest ones were seen for the C-cap (Table 3). The dArmRP molecules within the asymmetric unit are nearly identical (RMSD of 0.3 Å for all Cα atoms from residue 13 to 248) but the peptides differ significantly (Figure 8B). Mutations introduced on the surface had no large impact on the overall structure. Thus, structures of CAR2 and CAR2.V1 superimpose with a RMSD of 1.1 Å (Cα of molecules of CAR2 superimposed on chain A of CAR2.V1). This relatively large RMSD is a consequence of small structural changes in loops between helices 2 and 3 that are propagated along the solenoid. The structure of CAR2.V1 deviates slightly less than CAR2 from the curvature of the original model (RMSD of about 1.3 Å based on the Cα atoms of four internal repeats of CAR0). CAR2.V1 can be described analogously as a rather tall (large rise) and thin (small radius) cylinder with a large angle between neighboring repeats and Cα(P/P+2)-distances of 7.5 ± 0.2 Å (Figure 2C).

Despite the enlarged Cα(P/P+2)-distance based on the protein curvature, peptides are bound along the designed binding surface in an antiparallel orientation, and are fixed by several bidentate hydrogen bonds to asparagine residues (N^{37}) on the surface (Figure 7B) as intended in the original design and observed in nArmRPs. Superposition of all molecules revealed a highly conserved backbone of the bound peptide. Conformational fluctuations of the peptide backbone and side chains were observed towards the C-termini of the peptide, thus the highly conserved binding of the side chain is observed only for Arg2, Arg4 and Arg6 in pocket P1' and for Arg7 and Arg9 in pocket P2', respectively. The

increased conformational space sampled toward the ends of the peptide is consistent with its increased temperature factor as well as in a reduced number (e.g. chain A and E) or non-ideal geometry (chain B and F) of bidentate hydrogens bonds formed by N^{37} to the backbone of the bound peptide (Figure 7B). The measurement of the Cα-distances of the bound peptide revealed Cα(P/P+2)-distances of 6.6 ± 0.2 Å (measured for the distances between Arg3, Arg5, Arg7 and Arg9 of the peptide). Thus, binding of the poly-arginine peptide to protein CAR2.V1 with its non-optimal curvature was only possible by reducing the number of repeats involved in modular binding.

Despite the deviation from the optimal binding mechanism, single engineered pockets do bind the peptide side chains. As expected from the design, binding pockets are located between neighboring internal repeats, and are only partially formed if capping repeats are involved. Accordingly, the introduced binding pockets P2' make interactions with the positively charged side chains of the peptide (Arg5, Arg7 and Arg9). The binding mechanism of pocket P2' is highly conserved (Figure 7C). The mutation $N^{30}E$ allows the formation of a salt bridge between E^{30}-OE1 or OE2 and the peptide side chain Arg-NH1 or NH2 (Arg5, Arg7 and Arg9; Arg3 is excluded because no negatively charged amino acid is located at position 30 in the C-cap). Although the conformation of the introduced W^{33} is more variable, the mutation $A^{33}W$ allows the formation of a cation-π interaction [35]. This is achieved with W^{33} either in an upright (e.g. Trp117) or flattened conformation (e.g. Trp159). Although in CAR2.V1 no binding pockets were designed at position P1' (required to bind Arg2, 4, 6 and 8), residues G^{41}*, N^{1} and S^{40} allow the formation of a four-hydrogen-bond network and fix the peptide to the surface. G^{41}*-O interacts with R^{4}-NH1, N^{1}-O interacts with R^{4}-NH2 and S^{40} makes two hydrogen bonds, namely S^{40}-O with R^{4}-NH2 and S^{40}-OG with R^{4}-NE (Figure 7D). Thus, the mutation of $A^{40}S$ seems to be beneficial for binding arginines in pockets P1'. The mutations $N^{1}D$ and $L^{4}T$ additionally used for grafting P1' pockets in construct V3 and V4 did not always increase the affinity for $(KR)_n$ or $(RR)_n$ compared to the constructs missing these two mutations (V1 and V2) (see Table 1). This indicates that the designed complete binding pocket P1' in CAR2.V3 and V4 and the observed hydrogen network in CAR2.V1 and V2 that is made by wt residues and $A^{40}S$ are expected to fix the peptide side chains with similar efficacy.

The comparison of apo- and holo-structures of dArmRPs revealed that binding to the poly-arginine peptide had only moderate effects on the overall protein curvature. While structural variations observed in CAR2 are still present in CAR2.V1, indicating that the overall curvature is similar and a

certain amount of flexibility is maintained in the scaffold, the variations in the complex structure are slightly smaller. Nonetheless, the overall curvature of CAR2.V1 did not adapt upon binding to the peptide, and thus the peptide did not induce a complete modular binding all along its length.

The holo-structure also explained the beneficial effect on affinity of the charge neutralization mutation $K^{29}Q$ in CAR2.V2. K^{29} in CAR2.V1 is located on the binding surface close to binding pocket P2' (see Figure 7C). Thus, the positive charge of K^{29} would reduce the charge-charge interaction between the ligand arginine residue and E^{30}. This effect is strengthened by the multiple appearance of this pocket in the repetitive binding molecule.

Discussion

To design an ArmRP with curvature geometry suitable for modular peptide binding, a computational approach based on the multi-repeat model from yeast importin-α repeat pair GH was applied. Introduction of six MD-based mutations within the internal repeat, and the replacement of the N-cap by the consensus-based Y_{III}-cap were necessary to improve thermodynamic stability and to obtain crystal structures of dArmRPs. The structural analysis revealed that dArmRPs deviate from the planned curvature that would be optimal for binding peptides. The Cα(P/P+2)-distance is about 8.2 Å, which is significantly more than the desired binding distance of 6.7-7.0 Å. Nonetheless, modification of the binding surface of dArmRP resulted in binders, with the ability to bind $(KR)_n$ and $(RR)_n$ peptides. Affinities as high as 2.2 nM were determined by fluorescence anisotropy (Table 1). Importantly, the crystal structure of the complex with $(RR)_5$ clearly shows that the side chains occupy the engineered cavities, and thereby exclude non-specific electrostatic interactions as the main mode of binding, also consistent with the 1:1 stoichiometry deduced from the fluorescence anisotropy binding curve (Figure 6E).

Computationally designed armadillo scaffold

The C_{PAF}-cap of CAR0 was the only part of the *in silico* design without further modifications of which a structure could be obtained, and therefore allows to estimate the precision of the design. A comparison of the C_{PAF}-cap of the Rosetta-model and its X-ray structures (CAR2 and CAR2.V1)

superimpose with a low RMSD of 1.1 Å (based on the Cα atoms), and highlights the accuracy of the design (Figure 9A). The largest difference between model and structures was observed in the loop connecting helices 2 and 3. This loop is rearranged and allowed H^{22} to bind into the groove between the last internal repeat and the C-cap. Thereby, H^{22} forms a hydrogen bond with the backbone of A^{24} and is involved in the interaction network of the hydrophobic core by forming contacts with residue L^{19}, N^{24} and V^{27} of the C-cap and E^{25} and I^{28} of the internal repeat (Figure 9B). Notably, the same loop conformation is observed between internal repeats, although H^{22} is replaced by D^{22} there. In addition, the C_{PAF} capping repeat has a nearly identical structure to the consensus-based A_{II}-capping repeat [10] (PDB ID: 4DB6) (Figure 9C). Although they share only 63% sequence identity and 68% sequence similarity, the RMSD is 0.8 Å (based on the Cα residues of both C-caps).

Stabilizing effect of the MD-based Dq internal repeat

The structural stability of the CAR0 model was analyzed by MD-simulations, and several mutations were introduced in the internal repeats (Dq) at positions of high RMSF. The mutations M^3Q, L^4I, V^8I, $E^{21}D$, $K^{28}I$ and $A^{32}V$ (in CAR1) indeed increased the overall stability of this protein as confirmed experimentally (Table 1).

As Armadillo repeats are largely defined by the conservation of hydrophobic core residues, and as the stability of the protein largely depends on the continuous hydrophobic core [36], the mutations in the core ((L^4I, V^8I, $K^{28}I$ and $A^{32}V$) and intermediate-region (M^3Q)) are expected to be essential for the overall increase in stability (Figure 10). Among them, the side chain of Q^3 is partially involved in the hydrophobic core, but its hydrophilic end is involved in three conserved hydrogen bonds (Figure 10B). Methionine at position 3 cannot form these hydrogen bonds, and additional clashes are found for all possible rotamers sampled by Rosetta (data not shown).

The largest change in side chain size was the mutation $A^{32}V$. Here, several alternative residues (Ala, Leu, and Cys) were suggested by Rosetta. Leu showed no measurable advantage over Ala (Supplementary Figure S5 and Supplementary Table ST1), and Cys was not further considered to avoid issues with disulfide bond formation. MD simulations with A^{32} revealed a small cavity which was filled with Val, and this packing was confirmed in the CAR2 structure. In contrast, Leu, when modeled into this position, would clash with several other side chains of the core (Figure 10A).

Analogously, Ile fits best at position 8, since it is involved in more van der Waals interactions compared to Val, and Leu would result in clashes (Figure 10C). The unfavorable effects of L^8 were tested experimentally, and indeed protein stability was decreased by about 0.5 M GdnHCl (midpoints of denaturation, data not shown).

In position 4, Leu was replaced by Ile. In the structures of CAR2, Ile fills the hydrophobic core cavity without clashes, in contrast to Leu (Figure 10D). Therefore, Ile residues at position 4 and 8 seem be essential for the overall compactness and the stability of the structure. At position 28, both residues (Lys and Ile) seem to be valid solutions to strengthen the protein core; however, Ile better fills the hydrophobic core without leaving any cavity.

Curvature deviation from optimal template

Structural analysis of CAR2 showed that the desired curvature was not obtained, and the corresponding Cα(P/P+2)-distances are too long to match the bound peptide over a longer distance. For an efficient comparison of ArmRPs with different curvatures, we characterized the overall superhelical curvature by four parameters (rise, radius, angle and Cα(P/P+2)-distance).The accuracy of the parameterization depends on how well the structure can be described with the applied symmetry, and several tests verified the validity of the parameterization: first, each parameterized structure from CAR2, consisting only of the backbone Cα-atoms of two neighboring repeat pairs, superimpose with real structure pairs with a low RMSD of 0.5 Å ± 0.1 Å. Second, the Cα(P/P+2) distance measured on the parameterized structures doesn't deviate more than 0.1 Å from distances measured directly on the experimentally determined structures. The latter control is only indirect, because, as described above, Cα(P/P+2)-distances were measured after modeling the peptide to the apo-structures. Nonetheless, according to these control measurements and the manual inspection of the structures, the curvatures can be characterized with the given parameters.

The dArmRP CAR2 structures thus form superhelices which can be described by a rather thin and tall cylinder, and they display average Cα(P/P+2)-distances of 8.2 ± 0.7 Å, compared to the distance of 6.4 Å in the model structure CAR0. Although the most likely cause for this deviation is inaccuracy in computational design, two other factors may have affected the curvature: first, several modifications had to be introduced in the internal repeat of the initial Rosetta design sequence to stabilize the

protein and second, the computationally designed N-terminal cap was replaced in CAR2 with the consensus-based Y_{III} cap, as only the latter allowed crystallization.

The impact of stabilizing mutations on the overall curvature of the protein is difficult to predict but computational analysis can provide insights. We observed in the MD simulations that the mutations introduced in CAR1 induced a change of the overall curvature towards more unfavorable parameters for modular peptide binding, resulting in the curvature also observed in the experimentally determined structures of CAR2. Accordingly, the Cα(P/P+2)-distance changed in the model of CAR1 from 6.4 Å to 7.6 Å during the simulations. We hypothesize that the larger side chains introduced when converting the D_{SPVA} to the Dq repeat (L^4I, V^8I and $A^{32}V$) forced the protein to stretch from short and wide towards a longer but thinner shape. Among the introduced mutations, especially V^{32} seems to be a candidate to induce a curvature change, similar to what has been expected in the Rosetta-designed model of D_{PAAL} with L^{32} (Supplementary Figure S3). However, at this stage these explanations remain largely speculative because structural information from the original CAR0 sequence is not available.

Stabilization effect of the Y_{III}-capping repeat with the Dq internal repeat

Terminal capping repeats have been shown to be essential for the stability of repeat proteins, demonstrated e.g. for DARPins [22] and consensus-based ArmRPs [9, 10]. For dArmRPs, the impact of the Y_{III}-capping repeat was additionally tested on the original, computationally designed protein CAR0 ($N_V(D_{SPVA})_4C_{PAF}$). In contrast to the stabilizing effect of Y_{III} observed in CAR2 (Table 1), the addition of Y_{III} slightly reduced the stability against chemical denaturation of $Y_{III}(D_{SPVA})_4C_{PAF}$ (Supplementary Figure S9). Accordingly, Y_{III} and Dq must have a favorable interface complementarity, which enhances the overall protein stability by 0.4 M GdnHCl in comparison to Y_{III} and D_{SPVA}. In contrast to CAR2, no crystals could be obtained for $Y_{III}(D_{SPVA})_4C_{PAF}$, and thus we cannot experimentally investigate the differences in interaction.

These results do not allow us to decide whether the capping repeat has an influence on the overall curvature or not. Nevertheless, several designed caps in consensus-based ArmRPs (Y_{III}, Y_{II} and A_{III} and A_{II}) [10], have shown that the overall curvature is not influenced by the capping repeats. So far, the only influence was observed for internal repeat pairs adjacent to a domain-swapped cap, as found in early structures of consensus-based ArmRPs [10, 23]. In the CAR2 and CAR2.V1 structures, comparison of the curvature parameters of repeat pairs CD with repeat pairs BC and DE (which are

adjacent to the N- and C-cap, respectively), did not show significant differences (Figure 2). If capping repeats really had an effect, it would be expected to be stronger in directly adjacent repeats.

Peptide Binding by dArmRPs

The surface of CAR2 was modified, guided by structures of nArmRPs in complex with target peptides, and with the knowledge from consensus ArmRP-$(KR)_n$ interactions [31]. The introduction of the binding pocket P2' of importin-α into each internal repeat resulted in strong and specific binding molecules of positively charged peptides. The additional peptide-binding pocket P1' (mutation N^1D and I^4T) increased affinity only in some instances. One explanation for their small effect is the presence of a hydrogen bond network — present already in CAR2 — which allows the fixation of the corresponding peptide side chains at similar positions as predicted for P1' pockets, which therefore cannot contribute more binding energy. Nonetheless, the placement of a negatively charged residue as an anchor point in the P1' pocket could still be a valid strategy to improve binding of a positively charged side chain. Accordingly, D^1 could be replaced by E^1 to gain more flexibility, or shifted to position 41, as found in the pocket P2 of the major binding site of importin-α.

Our finding that the crystal structures of CAR2 and CAR2.V1 deviated from the expected modular binding scaffold is also reflected in the observed binding mode of the peptide. In the complex structure, dArmRP can bind a peptide, but only over a short stretch of approximately three repeats. Consequently, more flexibility is observed in the bound peptides towards their termini, indicating the loss of binding at the peptide ends.

In future design cycles, the structure should be further optimized to result in a curvature that will allow modular binding over more repeats. With the availability of an increasing number of crystal structures of designed ArmRPs, with or without bound peptide, more details about curvature and its impact on peptide binding will emerge. For example, we observed that CAR2 has larger structural fluctuations, as deduced from the larger standard deviations in their curvature parameters. Binding of peptide in CAR2.V1 seems to restrict the curvature to a state slightly closer to the desired one (reduced Cα(P/P+2)-distance).

In the future it needs to be investigated whether directed evolution experiments are capable of identifying mutations that alter the curvature to become compatible with modular binding over longer distance. Furthermore, knowledge from additional experimental structures of designed ArmRP, in

combination with protocols for enhanced sampling by MD [37], will likely allow a manipulation of curvature accurate enough to obtain modular binding over wide distances.

Materials and Methods

General molecular biology methods

Unless stated otherwise, experiments were performed as described previously [5].

Gene assembly and protein expression

Full-length gene assembly for proteins containing an N-terminal capping repeat, several internal repeats and a C-terminal capping repeat, were performed as described previously [5] using a step-by-step ligation of the modules digested by *Bsa*I or *Bpi*I or by a single multi-fragment ligation step (described below), similar to what has been described elsewhere [38, 39]. Exchange of capping repeats on full-length-constructs was performed by PCR amplification using two overhang fragments (20 bp) encoding the new capping repeat and the template fragment, which is lacking the corresponding capping repeat. Full-length genes for binder CAR.V1-CAR.V4 were synthesized by GeneArt® (Life Technologies) and cloned with *Bam*HI and *Hind*III into the pQE-derived vector.

Expression and purification for biophysical characterization as well as for crystallization was done as described previously [23]. Protein concentrations were determined by absorbance at 235 and 280 nm using molecular masses and extinction coefficients calculated with the tools available at the ExPASy proteomics server. Protein size and purity was assessed by 15% SDS-PAGE stained with Coomassie PhastGel Blue R-350 (GE Healthcare, Switzerland) and confirmed by mass spectroscopy.

Multi-fragment ligation assembly

Multi-fragment ligation (Supplementary Figure S4) was used to assemble dArmRPs with 1-10 identical internal repeats in one step. For this purpose, single modules were amplified by PCR with primer pairs pQE30_for + pQE30_short_KpnI_r, pQE30_short_BamHI_f + pQE30_short_KpnI_r or pQE30_short_BamHI_f + pQE30_rev from vectors containing the N-capping-, internal-, or C-capping repeat, respectively (primers are given in Supplementary Table ST3). The extended overhang sequence upstream and downstream of the N- and C-capping repeat provides an important quality

control to obtain full-length constructs after PCR amplification of the gel-purified multi-fragment ligation product. The corresponding PCR fragments were digested by *Bsa*I, *Bsa*I+*Bpi*I and *Bpi*I, and purified by NucleoSpin columns (Macherey-Nagel). Ligation was performed in three steps: First, 0.5 µg of single-digested N-cap fragments were ligated with a 5-fold excess of double-digested internal repeat fragments using 2.5 U of T4 DNA Ligase and incubated for 15' at room temperature. In a second step, 0.5 µg of digested C-cap fragment and 1 U of fresh ligase were added to the mixture after buffer adjustment according to the volume increase and incubated for 30'. In a last step (optional), 0.5 µg of N- and C-cap fragments were added again to the mixture to increase the amount of full-length constructs. The buffer was adjusted according to volume increase and incubated for another 15'. To obtain constructs with the right number of internal repeats, the ligation mixture was heat-inactivated at 65°C for 10' prior to loading and separated on a 1.5% agarose gel. The desired DNA bands were extracted and amplified by PCR with outer primers pQE30_for+pQE30_rev using 50 ng template DNA. Full-length fragments were inserted into cloning and expression vectors pQE30 or pPANK by *Bam*HI and *Hind*III sites. Proper assembly of constructs was validated by colony-PCR and DNA sequencing.

Model backbone template generation

A backbone model based only on the GH curvature from importin-α (PDB ID: 1EE4 [17]) was generated by iterative superposition of the repeat G from the GH repeat pair on the H repeat of a copy of the GH repeat (Supplementary Figure S2). The GH-backbone template was then used to repack the protein by the Rosetta Program.

In silico protein design (Rosetta software)

Symmetric constraints were applied throughout the design trajectories. A single 'master' ArmRP domain was designated arbitrarily and all side chain packing, minimization, and backbone minimization moves were done simultaneously on this master domain and all other domains in the system. Each move consisted of combinatorial side chain design and conformational search, backbone, and side chain minimization. An extended peptide, corresponding to the target peptide for binding, was maintained throughout the simulations in the preferred orientation observed in the yeast importin-α crystal structure (PDB 1EE4 [17]). The Asn residues at position 37 that contact the peptide backbone were maintained in their native orientations and the rigid-body orientation of the peptide-ArmRP complex was minimized during the simulation. Resulting models were analyzed for their

peptide-backbone binding ability, and structural integrity. Further symmetric substitutions were introduced on the surface of the ArmRP by adding salt bridges and polar groups to increase solubility.

Analysis of superhelical parameters

Superhelical parameters were determined by analyzing the geometry of internal repeat pairs using the generalized helix description as it has been implemented in the *make_symmdef_file.pl* script from the Rosetta symmetry framework [18]. As input structures we used the Cα-atom coordinates from 41 residues of two consecutive internal repeats (the flexible residues at position 23 were excluded). Curvature parameters as depicted in Figure 2 were first generated for each pair of internal repeats (M_1:M_2, M_2:M_3, M_3:M_4, and M_4:M_5) and for each molecule found within the asymmetric unit and then averaged (Supplementary Figure S1). The angle 2·Ω (°) describes the angle between the centers of mass of two consecutive internal repeats and the central helix axis (Supplementary Figure S1C).

The Cα(P/P+2)-distance was determined for each symmetrized repeat pair model. Thereby a fragmented peptide is modeled to the multi-repeat model by superposition of repeat D of the designed ArmRP binding to the (KRK)-peptide fragment [31] on each repeat of the multi-repeat model. Based on the artificial peptide along the whole binding surface of the multi-repeat models, the Cα(P/P+2)-distances (Supplementary Figure S1D) were measured and correlated with the parameter set of the multi-repeat model.

SEC and MALS

Size exclusion chromatography (SEC) coupled with multi-angle light scattering measurements (MALS) were carried out using a liquid chromatography system (Agilent LC1100, Agilent Technologies, Santa Clara, CA) coupled to an Optilab rEX refractometer (Wyatt Technology, Santa Barbara, CA) and a miniDAWN three-angle light-scattering detector (Wyatt Technology). For protein separation, 50 µl of 50 µM designed ArmRP were injected on a Superdex 200 10/30 column (GE Healthcare) and run at 0.5 ml/min in buffer 50 mM Tris, 150 mM NaCl, pH 7.6. Analysis of the data was done with the ASTRA software (version 5.2.3.15; Wyatt Technology).

ANS binding

The fluorophore 1-anilino-naphthalene-8-sulfonate (ANS) binds to exposed hydrophobic patches or pockets in proteins, whereupon its fluorescence increases significantly. In this study, ANS

fluorescence was used to probe the packing of the designed hydrophobic cores. The measurements were performed at 20°C by adding ANS (final concentration 100 µM) to 10 µM of purified protein in 50 mM Tris, 150 mM NaCl, pH 7.6. Unless stated otherwise, the fluorescence signal was recorded using 600 µl of protein sample in a quartz cuvette (light path 5 mm) and a FluoroMax®-4 spectrofluorometer (Horiba Scientific). The emission spectrum from 400 to 650 nm (bandwidth, 2 nm, data pitch, 1 nm; integration time, 0.1 s) was recorded (as CPS) with an excitation wavelength of 350 nm (bandwidth, 2 nm). For each sample, three spectra were recorded and averaged and the blank value subtracted.

CD spectroscopy and unfolding curves

CD measurements were performed on a Jasco J-810 spectropolarimeter (Jasco, Japan) using a 10 µM protein solution (in 50 mM Tris, pH 7.6 and 150 mM NaCl) in a 0.5-mm cylindrical thermo-cuvette. CD spectra were recorded from 190 to 250 nm with a data pitch of 0.5 nm, a scan speed of 100 nm/min, a response time of 4 s, and a bandwidth of 1 nm. Each spectrum was recorded four times and averaged. Measurements were performed at 20°C. The CD signal was blank-corrected and converted to MRE. GdnHCl-induced denaturation measurements were performed after overnight incubation at 20°C with increasing concentrations of GdnHCl (99.5% purity, Fluka), and the data were collected at 222 nm (data pitch, 1 s; response time, 4 s; bandwidth, 2 nm; measured time, 45 s) and processed as described above. Heat denaturation curves were obtained by measuring the CD signal at 222 nm with temperature increasing from 20 to 92°C using an external water bath (Julabo FS18) (data pitch, 0.2°C; heating rate, 1°C/min; response time, 4 s; bandwidth, 1 nm).

GdnHCl-induced denaturation data showed slopes of the pretransition and posttransition phases either close to zero or not well defined, and thus they were set to zero. Data were thus fitted to a two-state unfolding model without sloping baselines (eq. 1). Fits were only used to estimate the unfolding midpoint and not other parameters.

$$Y = \frac{y_f + y_d \times e^{-\frac{\Delta G}{RT}}}{1 + e^{\frac{-\Delta G}{RT}}} \qquad \text{(eq. 1)}$$

with $\Delta G = \Delta Go - m[GdnHCl]$, where Y is the fraction of unfolded protein (expressed as normalized MRE); y_f and y_d are the signals for fully folded and fully denatured proteins, respectively; R is the universal gas constant, T is the temperature (298 K), [GdnHCl] is the concentration of GdnHCl and $[GdnHCl]_{1/2}$ is the GdnHCl concentration at the unfolding midpoint, which is obtained from

$[\mathrm{GdnHCl}]_{1/2} = \frac{\Delta Go}{m}$. For the denaturation plots, values of MRE (blank corrected), were normalized by setting the pretransition values (folded) as 0 and the putative completely unfolded protein (defined as MRE=0) as 1.

Thermal denaturation curves were fitted to a two-state unfolding model with sloping baselines according to Jackson and Fersht [40] with eq. 2 and 3. Since the full reversibility and two-state nature of this system is questionable, all fits were only used to estimate the midpoint of thermal denaturation and not other parameters:

$$\Delta G = \frac{T_m - T}{T_m} \times \Delta H - (T_m - T) \times \Delta C_p + T \times \Delta C_p \times ln \frac{T_m}{T} \quad \text{(eq. 2)}$$

$$MRE = \frac{1}{1+\frac{1}{e^{\frac{\Delta G}{RT}}}} \times (y_l + m_l \times T - y_u - m_u \times T) + y_u + m_u \times T \quad \text{(eq. 3)}$$

Here y_l and y_u are the y-axis intercepts of the lower and upper baseline, respectively; m_l and m_u are the slopes of the lower and upper baseline, respectively; T is the temperature; T_m the midpoint of thermal denaturation; ΔG and ΔH are the free energy and enthalpy of unfolding, respectively, and ΔC_p is the change in heat capacity at constant pressure.

NMR spectroscopy

All 1D ^{1}H-NMR spectra were recorded at 37°C using 0.5 mM protein solutions in 50 mM phosphate buffer pH 7.6, 30 mM NaCl. One-dimensional proton NMR spectra were recorded with water suppression by standard presaturation. To obtain proton-nitrogen correlation maps standard [^{15}N,^{1}H]-HSQC spectra were recorded on Bruker Avance 700 MHz NMR spectrometer with uniformly ^{15}N-labeled labeled protein in 50 mM phosphate buffer pH 7.6, 150 mM NaCl [41]. Data were processed and inspected in Topspin 2.1.

Crystallization and structure determination

Sparse-matrix screens from Hampton Research (California) and Molecular Dimensions (Suffolk, UK) were used to identify the preliminary crystallization conditions in 96-well Corning plates (Corning Incorporated, New York) at 4°C. A Phoenix crystallization robot (Art Robbins Instruments) was used to perform sitting drop vapor-diffusion experiments. The protein solutions were filtered through a 0.22 µm Milex® filter (Millipore). Prior to crystallization, protein CAR2.V1 was supplemented with a 1.5-fold molar excess of $(RR)_5$ peptide (the peptide was dissolved in water and changed the volume of the

sample by 1%). Proteins CAR2 and CAR2.V1_nohis were mixed with reservoir solutions at 1:1, 1:2, or 2:1 ratios (200 – 300 nl final volume) and at 1:1, 1:2 or 1:5 (300 nl), respectively, and the mixtures were equilibrated against 50 μl of reservoir solution at 4°C. Table 2 summarizes all the crystallization conditions, data collection and refinement statistics. Crystals for CAR2 and CAR2.V1 were washed three times in reservoir solution supplemented with 10% glycerol and 20% ethylene glycol, respectively, before being flash-cooled in liquid nitrogen. Using a Pilatus detector system on beam line X06DA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland), data were collected and processed using program XDS [42]. The CAR2 structure was solved by molecular replacement (software PHASER [43]) using a structure of another CAR2 variant which had been determined by a poly-alanine search model that was created from the crystal structure of a consensus ArmRP (chain A from PDB ID: 4V3R [44]). CAR2.V1_nohis was determined by using CAR2 as model. Refinement was done using programs PHENIX-Refine [45], REFMAC5 [46] and COOT [47]. Water molecules were added to well-defined difference electron density peaks at H-bond distance from the protein. $(RR)_5$ peptides were identified in the final electron density maps of CAR2.V1. The program PROCHECK [48] was used to validate the final structures, and PyMOL was used to generate figures [49].

Diffraction data of CAR2.V1 suggested an orthorhombic space group, and structure determination was possible in space group $P2_12_12_1$ with two molecules related by a two-fold axis. In this setting, the non-crystallographic rotation axis was almost parallel to the c-axis. However, the refinement did not converge and the R_{free} value never dropped below 29.3%. Therefore, the symmetry restraints were relaxed and the structure was refined in space-group $P2_1$ with four molecules in the asymmetric unit. All four chains show small but significant deviations in ligand binding, which confirms that the assignment of a monoclinic space group is correct, although the crystal lattice is almost orthorhombic. Further analysis of the diffraction data using a L-test [50] implemented in the program TRUNCATE [51] suggested pseudo-merohedral twinning with a twinning fraction of 49.5%. Only after taking twinning into account refinement converged at final R_{cryst} and R_{free} values of 20.19% and 25.53%, respectively.

Molecular dynamics (MD) simulations

Explicit solvent MD simulations were performed at constant temperature (310 K) and constant pressure (1 atm) using the v-rescale thermostat [52] and Berendsen pressure coupling [53]. The long-range electrostatic interactions were treated by the particle mesh Ewald method [54]. The van-der-Waals interactions were truncated at a cut-off of 9 Å. The LINCS algorithm [55] was used to fix the

length of all bonds. Virtual sites were used for removing the fastest degrees of freedom, which allowed an integration time step of 5 fs. Structures were saved every 5 ps for analysis. The MD simulations were carried out using the Gromacs program [56] with the OPLS force field [57, 58] and TIP3P potential [59] for water molecules. The first 200 ns of each MD run were considered as equilibration time and were excluded from the RMSF calculation. The block average time in RMSF calculation was 2 ns. All the constructs were simulated in the absence of the His tag.

ELISA

A MaxiSorp plate (Nunc) was coated with NeutrAvidin (100 µl, 66 nM in PBS, overnight at 4°C) and then blocked with PBS-TB (300 µl, 0.1% Tween and 0.2% BSA in PBS [60], 1 h at room temperature). The target peptides (expressed as pD-fusion [5, 29] or chemically synthesized (JPT) (Supplementary Table ST4)) were immobilized via their biotin residues on NeutrAvidin (100 µl, 200 nM in PBS-TB). Buffers for binding and washing in all ELISA experiments were PBS-TB and PBS-T (300 µl, 0.1% Tween in PBS), respectively. Purified proteins (100 µl, 200 nM in PBS-TB) were incubated with the target for 1 h at 4°C. Wells were washed three times with 300 µl of PBS-T before detection of the proteins with a primary anti-$RGSH_6$ antibody (100 µl, 1:5000 dilution in PBS-TB, 45 min at 4°C; Qiagen, Germany) and a secondary goat anti-mouse IgG alkaline phosphatase conjugate antibody (100 µl, 1:10,000 in PBS-TB, 45 min at 4°C; Sigma). Absorbance was measured at 405 nm (and 540 nm reference wavelength) using a Tecan Infinite M1000 plate reader after incubation with the substrate disodium 4-nitrophenyl phosphate (100 µl, 3 mM in buffer containing 50 mM $NaHCO_3$ and 50 mM $MgCl_2$, 60 min at 37°C; Fluka).

Anisotropy measurement

The assays were performed in black non-binding 96-well plates (Greiner). 2 nM or 10 nM of peptide-sfGFP fusion protein was titrated with increasing concentrations of ArmRP. The concentration of peptide-sfGFP was chosen to be maximally two-fold over the respective K_d (CAR.V2), but preferentially below K_d (CAR.V1/V3/V4 and CAR2). For the variants, a dilution series of 24 concentrations of dArmRP was used and fluorescence anisotropy was measured on a Safire II plate reader (Tecan). Data were averaged from four samples and the anisotropy value from the lowest ArmRP concentration was subtracted for normalization. Data were fitted by a simple one-to-one binding model using SigmaPlot®, using eq. 4:

$$[AB] = \left(-\frac{1}{2}\right) \cdot \left[-(K_\mathrm{d} + [A_{tot}] + [B_{tot}]) + \sqrt{(K_\mathrm{d} + [A_{tot}] + [B_{tot}])^2 - 4 \cdot [A_{tot}] \cdot [B_{tot}]}\right] \qquad \text{(eq. 4)}$$

where [*AB*] is the complex formed, [A_{tot}] and [B_{tot}] are the concentrations of peptide and ArmRP, respectively, and K_d is the dissociation constant, and $\frac{[AB]}{[A_{tot}]} = \frac{(P - P_{min})}{(P_{max} - P_{min})}$, where *P* is the measured anisotropy, P_{min} its minimum in the absence of ArmRP and P_{max} the value when the peptide is fully bound by protein.

Acknowledgements

We would like to thank Céline Stutz-Ducommun and Beat Blattmann from the high-throughput crystallization center and the staff from beamlines X06SA and X06DA from the Swiss Light Source for skillful technical support. Additional thanks goes to Dr. Johannes Schilling for helpful discussions. SJF was supported by the Human Frontier Science Program. This work was financially supported by a Swiss National Science Foundation grant (Sinergia S-41105-06-01).

References

[1] F. Diella, N. Haslam, C. Chica, A. Budd, S. Michael, N.P. Brown, et al., Understanding eukaryotic linear motifs and their role in cell signaling and regulation, Front. Biosci. 13 (2008) 6580-6603.
[2] T. Pawson, P. Nash, Assembly of cell regulatory systems through protein interaction domains, Science 300 (2003) 445-452.
[3] E. Petsalaki, A. Stark, E. Garcia-Urdiales, R.B. Russell, Accurate prediction of peptide binding sites on protein surfaces, PLoS Comput. Biol. 5 (2009) e1000335.
[4] C. Reichen, S. Hansen, A. Plückthun, Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins, J. Struct. Biol. 185 (2014) 147-162.
[5] F. Parmeggiani, R. Pellarin, A.P. Larsen, G. Varadamsetty, M.T. Stumpp, O. Zerbe, et al., Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core, J. Mol. Biol. 376 (2008) 1282-1304.
[6] R. Aasland, C. Abrams, C. Ampe, L.J. Ball, M.T. Bedford, G. Cesareni, et al., Normalization of nomenclature for peptide motifs as ligands of modular protein domains, FEBS Lett. 513 (2002) 141-144.
[7] M. Hatzfeld, The armadillo family of structural proteins, Int. Rev. Cytol. 186 (1999) 179-224.
[8] P. Forrer, H.K. Binz, M.T. Stumpp, A. Plückthun, Consensus design of repeat proteins, ChemBioChem 5 (2004) 183-189.
[9] P. Alfarano, G. Varadamsetty, C. Ewald, F. Parmeggiani, R. Pellarin, O. Zerbe, et al., Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy, Protein Sci. 21 (2012) 1298-1314.
[10] C. Madhurantakam, G. Varadamsetty, M.G. Grütter, A. Plückthun, P.R. Mittl, Structure-based optimization of designed Armadillo-repeat proteins, Protein Sci. 21 (2012) 1015-1028.

[11] R. Das, D. Baker, Macromolecular modeling with Rosetta, Annu. Rev. Biochem 77 (2008) 363-382.
[12] K. Park, B.W. Shen, F. Parmeggiani, P.S. Huang, B.L. Stoddard, D. Baker, Control of repeat-protein curvature by computational protein design, Nat. Struct. Mol. Biol. 22 (2015) 167-174.
[13] F. Parmeggiani, P.S. Huang, S. Vorobiev, R. Xiao, K. Park, S. Caprari, et al., A General Computational Approach for Repeat Protein Design, J. Mol. Biol. 427 (2015) 563-575.
[14] T.J. Brunette, F. Parmeggiani, P.S. Huang, G.B. Habha, D.C. Ekiert, S.E. Tsutakawa, et al., Exploring the repeat protein universe through computational protein design, Nature 528 (2015) 580-584.
[15] L. Doyle, J. Hallinan, J. Bolduc, F. Parmeggiani, D. Baker, B.L. Stoddard, et al., Rational design of alpha-helical tandem repeat proteins with closed architectures, Nature 528 (2015) 585-588.
[16] E. Conti, J. Kuriyan, Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha, Structure 8 (2000) 329-338.
[17] E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha, Cell 94 (1998) 193-204.
[18] F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, I. André, Modeling symmetric macromolecular structures in Rosetta3, PLoS One 6 (2011) e20450.
[19] N.P. King, W. Sheffler, M.R. Sawaya, B.S. Vollmar, J.P. Sumida, I. Andre, et al., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy, Science 336 (2012) 1171-1174.
[20] R. Das, I. Andre, Y. Shen, Y.B. Wu, A. Lemak, S. Bansal, et al., Simultaneous prediction of protein folding and docking at high resolution, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 18978-18983.
[21] T. Spreter, C.K. Yip, S. Sanowar, I. Andre, T.G. Kimbrough, M. Vuckovic, et al., A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system, Nat. Struct. Mol. Biol. 16 (2009) 468-476.
[22] S.K. Wetzel, G. Settanni, M. Kenig, H.K. Binz, A. Plückthun, Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins, J. Mol. Biol. 376 (2008) 241-257.
[23] C. Reichen, C. Madhurantakam, A. Plückthun, P.R. Mittl, Crystal structures of designed armadillo repeat proteins: Implications of construct design and crystallization conditions on overall structure, Protein Sci. 23 (2014) 1572-1583.
[24] L. Lüthy, M.G. Grütter, P.R.E. Mittl, The crystal structure of Helicobacter pylori cysteine-rich protein B reveals a novel fold for a penicillin-binding protein, J. Biol. Chem. 277 (2002) 10187-10193.
[25] T. Merz, S.K. Wetzel, S. Firbank, A. Plückthun, M.G. Grütter, P.R.E. Mittl, Stabilizing ionic interactions in a full-consensus ankyrin repeat protein, J. Mol. Biol. 376 (2008) 232-240.
[26] M.A. Kramer, S.K. Wetzel, A. Plückthun, P.R. Mittl, M.G. Grütter, Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module, J. Mol. Biol. 404 (2010) 381-391.
[27] B. Catimel, T. Teh, M.R. Fontes, I.G. Jennings, D.A. Jans, G.J. Howlett, et al., Biophysical characterization of interactions involving importin-alpha during nuclear import, J. Biol. Chem. 276 (2001) 34189-34198.
[28] M.R. Hodel, A.H. Corbett, A.E. Hodel, Dissection of a nuclear localization signal, J. Biol. Chem. 276 (2001) 1317-1325.
[29] G. Varadamsetty, D. Tremmel, S. Hansen, F. Parmeggiani, A. Plückthun, Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity, J. Mol. Biol. 424 (2012) 68-87.
[30] C. Ewald, M.T. Christen, R.P. Watson, M. Mihajlovic, T. Zhou, A. Honegger, et al., A combined NMR and computational approach to investigate Peptide binding to a designed armadillo repeat protein, J. Mol. Biol. 427 (2015) 1916-1933.
[31] S. Hansen, D. Tremmel, C. Madhurantakam, C. Reichen, P.R. Mittl, A. Plückthun, Structure and Energetic Contributions of a Designed Modular Peptide-Binding Protein with Picomolar Affinity, J. Am. Chem. Soc. 138 (2016) 3526-3532.
[32] A.L. Cortajarena, F. Yi, L. Regan, Designed TPR modules as novel anticancer agents, ACS Chem. Biol. 3 (2008) 161-166.

[33] A.L. Cortajarena, T. Kajander, W.L. Pan, M.J. Cocco, L. Regan, Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins, Protein Eng. Des. Sel. 17 (2004) 399-409.
[34] N. Sawyer, B.M. Gassaway, A.D. Haimovich, F.J. Isaacs, J. Rinehart, L. Regan, Designed Phosphoprotein Recognition in Escherichia coli, ACS Chem. Biol. 9 (2014) 2502-2507.
[35] D.A. Dougherty, Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp, Science 271 (1996) 163-168.
[36] A.H. Huber, W.J. Nelson, W.I. Weis, Three-dimensional structure of the armadillo repeat region of beta-catenin, Cell 90 (1997) 871-882.
[37] M. Bacci, A. Vitalis, A. Caflisch, A molecular simulation protocol to avoid sampling redundancy and discover new states, Biochim. Biophys. Acta, Gen. Subj. 1850 (2015) 889-902.
[38] A. Urvoas, A. Guellouz, M. Valerio-Lepiniec, M. Graille, D. Durand, D.C. Desravines, et al., Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats, J. Mol. Biol. 404 (2010) 307-327.
[39] M. Nikkhah, Z. Jawad-Alami, M. Demydchuk, D. Ribbons, M. Paoli, Engineering of beta-propeller protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat, Biomol. Eng 23 (2006) 185-194.
[40] S.E. Jackson, A.R. Fersht, Folding of Chymotrypsin Inhibitor-2 .1. Evidence for a 2-State Transition, Biochemistry 30 (1991) 10428-10435.
[41] R.P. Watson, M.T. Christen, C. Ewald, F. Bumbak, C. Reichen, M. Mihajlovic, et al., Spontaneous self-assembly of engineered armadillo repeat protein fragments into a folded structure, Structure 22 (2014) 985-995.
[42] W. Kabsch, XDS, Acta Crystallogr. Sect. D. Biol. Crystallogr. 66 (2010) 125-132.
[43] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658-674.
[44] C. Reichen, C. Madhurantakam, S. Hansen, M.G. Grütter, A. Plückthun, P.R.E. Mittl, Structures of designed armadillo-repeat proteins show propagation of inter-repeat interface effects, Acta Crystallogr. Sect. D. Biol. Crystallogr. 72 (2016) 168-175.
[45] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. Sect. D. Biol. Crystallogr. 66 (2010) 213-221.
[46] G.N. Murshudov, A.A. Vagin, A. Lebedev, K.S. Wilson, E.J. Dodson, Efficient anisotropic refinement of macromolecular structures using FFT, Acta Crystallogr. Sect. D. Biol. Crystallogr. 55 (1999) 247-255.
[47] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. Sect. D. Biol. Crystallogr. 60 (2004) 2126-2132.
[48] R.A. Laskowski, D.S. Moss, J.M. Thornton, Main-Chain Bond Lengths and Bond Angles in Protein Structures, J. Mol. Biol. 231 (1993) 1049-1067.
[49] W.L. DeLano. PyMOL. 2002.
[50] J.E. Padilla, T.O. Yeates, A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning, Acta Crystallogr. Sect. D. Biol. Crystallogr. 59 (2003) 1124-1130.
[51] S. French, K. Wilson, Treatment of Negative Intensity Observations, Acta Crystallogr. Sect. A: Found. Crystallogr. 34 (1978) 517-525.
[52] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. 126 (2007).
[53] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Dinola, J.R. Haak, Molecular-Dynamics with Coupling to an External Bath, J. Chem. Phys. 81 (1984) 3684-3690.
[54] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, J. Chem. Phys. 98 (1993) 10089-10092.
[55] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: A linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463-1472.

[56] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation, J. Chem. Theory Comput. 4 (2008) 435-447.
[57] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, J. Phys. Chem. B 105 (2001) 6474-6487.
[58] W.L. Jorgensen, D.S. Maxwell, J. TiradoRives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225-11236.
[59] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, J. Chem. Phys. 79 (1983) 926-935.
[60] J. Sambrook, D.W. Russell. Molecular cloning : a laboratory manual. 3rd. ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2001.

Tables

Table 1. Biophysical properties of dArmRPs

Constructs	Short name	Residues [b]	pI[c]	MW_{cal} (kDa)[d]	MW_{obs} (kDa)[e]	OS[f]	$MW_{obs/cal}$[g]	CD_{222} (MRE)[h]	T_m (°C)[i]	CD_{GdnHCl} (M)[j]	K_d (nM)[k]			
											$(RR)_5$	$(RR)_4$	$(KR)_5$	$(KR)_4$
$N_V(D_{SPVA})_4C_{PAF}$	CAR0	251	5.3	26.3	30.4	mono	1	-15,121	93 ± 1.8	2.2	n.d.	n.d.	n.d.	n.d.
$N_V(Dq)_4C_{PAF}$	CAR1	251	5	26.3	31.9	mono	0.9	-16,824	73 ± 0.2	2.8	n.d.	n.d.	n.d.	n.d.
$Y_{III}(Dq)_4C_{PAF}$	CAR2	251	5.1	26.3	31.9	mono	0.9	-14,997	73 ± 0.5	3	n.d.	n.d.	$>1\cdot10^4$	n.d.
$Y_{III}(Dq)_4C_{PAF}$	CAR2_nohis	243	4.7	25.2	35.3	mono	0.9	-13,422	73 ± 0.3	3	$>1\cdot10^4$	$>1\cdot10^4$	$>1\cdot10^4$	$>1\cdot10^4$
$Y_{III}(Dq.V1)_4Cq$	CAR2.V1_nohis	243	4.6	25.8	35.1	mono	0.9	-16,861	64 ± 0.4	2.7	24 ± 4	112 ± 9	76 ± 6	860 ± 120
$Y_{III}(Dq.V2)_4Cq$	CAR2.V2_nohis	243	4.4	25.8	36.8	mono	1	-14,486	50 ± 0.8	2.5	2.2 ± 0.1	10 ± 1	8.0 ± 1	134 ± 2
$Y_{III}(Dq.V3)_4Cq$	CAR2.V3_nohis	243	4.5	25.8	39.9	mono	0.9	-14,735	56 ± 0.8	1.8	44 ± 4	187 ± 2	24 ± 2	331 ± 19
$Y_{III}(Dq.V4)_4Cq$	CAR2.V4_nohis	243	4.3	25.8	60.5	mono	0.9	-13,606	51 ± 10.7	1.6	60 ± 3	348 ± 8	32 ± 3	477 ± 2
$Y_IM_4A_I$ [l]		253	4.7	27.1	34.4	mono	1	-15,474	73 ± 0.5	3.4	n.d.	n.d.	n.d.	n.d.
$Y_{III}M_4A_{II}$ [l]		252	4.8	26.8	56.9	mix[e]	1.4	-17,056	87 ± 1.5	3.9	n.d.	n.d.	36 ± 1[m]	265 ± 23[m]

[a] dArmR: N-cap (e.g., N_V and Y_{III} [k]), C-cap (e.g C_{PAF} and A_{II} [k]), and internal repeats (e.g. D_{SPVA}, Dq or Dq.V1).

[b] The number of residues includes the $MRGSH_6GS$ tag; all constructs consist of six repeats including capping repeats.

[c] Isoelectric point (pI), calculated from the sequence.

[d] Molecular weight calculated from the sequence.

[e] Observed molecular weight as determined by SEC.

[f] Oligomeric state (OS) measured by multi-angle static light scattering (MALS). Mono: monomeric state; Mix: equilibrium between monomer and dimer.

[g] Ratio between observed (by MALS) and molecular weight calculated from the sequence ($MW_{obs/calc}$).

[h] Mean residue ellipticity at 222 nm expressed as deg·cm^2/dmol.

[i] Transition midpoint (T_m) observed in thermal denaturation measured by CD. Approximations.

[j] Midpoint of transition in GdnHCl-induced denaturation, measured by CD.

[k] Equilibrium dissociation constant against (peptide)-sfGFP determined by fluorescence anisotropy.

[l] Consensus-based ArmRP. M refers to the consensus-based internal repeat $\overline{M}$ reported by Alfarano et al. [13]

[m] Data from Hansen et al. [35]

n.d.: not determined

Table 2. Data and refinement statistics

	$Y_{III}(Dq)_4C_{PAF}$= CAR2	$Y_{III}(Dq.V1)_4C_{PAF}$= CAR2.V1
PDB ID	**4D4E**	**4D49**
Structure	with $MRGSH_6$-tag no peptide	without $MRGSH_6$-tag with $(RR)_5$ peptide
Crystallization condition	0.3 M Na-acetate pH 5.5 0.1 M Na-acetate 25% PEG 2K MME	0.1 M Tris/HCl pH 8.5 0.2 M Mg-chloride 10% PEG 1000 10% PEG 8000
Data statistics		
Cell parameters:	a: 51.81 Å, b: 68.67 Å, c:126.84 Å α: 90.0°, β: 90.0°, γ: 90.0°	a: 88.57 Å, b: 51.73 Å, c:107.61 Å α: 90.0°, β: 90.16°, γ: 90.0°
Space group:	$P2_12_12_1$	$P2_1$
Resolution range:	39.35 – 2.0 Å (2.11-2.0)	20.0 – 2.1 Å (2.15-2.1)
Number of molecules/ AU	2	4
Number of reflections		
Observed	139,251 (16,131)	241,736 (18,790)
Unique	27,662 (3,234)	106,755 (8,086)
I/σ (I)	9.3 (2.3)	5.01 (1.37)
Completeness (%)	88.2 (72.3)	95.7 (97.4)
[1] R_{merge} (%)	8.6 (58.8)	16.7 (150.9)
Multiplicity	5.0 (5.0)	2.26 (2.32)
Refinement statistics		
[2] R_{cryst} (%)	18.75	20.09
[3] R_{free} (%)	26.34	25.45
Number of protein atoms	3,490	7552
Number of waters	250	238
Number of hetero atoms	18	22
Number of chains:	2	4
RMSD from ideal geometry		
Bond length (Å)	0.016	0.015
Bond angles (°)	1.875	1.795
B values ($Å^2$)		
Wilson B	31.8	21.6
Average B	33.5	43.5
Ramachandran Plot (%)		
Residues in preferred regions	97.85	98.56
Residues allowed regions	1.94	1.44
Residues in disallowed regions	0.22	0.00

Values in parentheses are for the highest resolution shell.

[2] $R_{cryst}=\Sigma|F_{obs}-F_{calc}|/\Sigma F_{obs}$, where F_{obs} and F_{calc} are the observed and the calculated structural factors, respectively

[3] R_{free} was calculated using 5 % of the reflections similar to R_{cryst}.

Table 3. Average B-factors of armadillo repeats (all values given in Å^2).

	chain	N-cap	Internal	C-cap
$Y_{III}(Dq)_4C_{PAF}$= CAR2	A	36.9	25.1	29.3
	B	30.0	29.1	55.3
$Y_{III}(Dq.V1)_4C_{PAF}$= CAR2.V1	A	51.1	44.2	58.5
	B	44.9	31.1	63.5
	E	35.7	30.1	49.4
	F	45.6	44.4	75.8

‡ B factors were calculated from backbone atoms.

Figure legends

Figure 1. Peptide distance analysis of natural ArmRPs of the importin-α family. A: Detailed view of the major binding site, composed of repeats B-E shown as grey cylinders for yeast importin-α (PDB ID 1BK6) with bound NLS peptide (green), making six backbone hydrogen bonds (yellow dashed lines) with the conserved Asn residues (orange). Interaction residues of importin-α with peptide side chains are shown in yellow. Ionic interactions are indicated by blue dashed lines. The Cα(P/P+2)-distance (red dashed lines) is measured between the Cα–atoms (red spheres) of the peptide residues bound by Asn^{37}. B: Summary of predicted Cα(P/P+2)-distances found in repeat pair models of importin-α (BC to HI) distinguished by organism (calculated as described in Supplementary Figure S1). N- and C-terminal repeats were excluded (repeat A and J). Upper and lower Cα(P/P+2)-distances needed for continuous modular binding are indicated by black dashed lines. (PDB ID: human: 2JDQ, 3FEX, 3FEY, 3TJ3; mouse: 1EJL, 1EJY, 1IAL, 1IQ1, 1PJM, 1PJN, 1Q1S, 1Q1T, 1Y2A, 2C1M, 3BTR, 3KND, 3L3Q, 3OQS, 3Q5U, 3RZ9, 3RZX, 3TPM, 3UKW, 3UKX, 3UKY, 3UKZ, 3UL0, 3UL1, 3UVU, 3VE6, 4HTV ; yeast : 1BK6, 1EE4, 1EE5, 1UN0, 2C1T).

Figure 2. Curvature parameterization of ArmRPs. Parameterization of each pair of neighboring internal repeats has been applied by using helical symmetry operation from Rosetta (described in Supplementary Figure S1). Multi-repeat models of each internal repeat pair (BC to HI) of A: importin-α (PDB ID 1EE4 [16]) and the model structure CAR0 ($N_v(D_{SPVA})_4C_{PAF}$). B: dArmRP CAR2 ($Y_{III}(Dq)_4C_{PAF}$), and C: dArmRP CAR2.V1 ($Y_{III}(Dq.V1)_4C_{PAF}$). Parameter values of repeat pairs denote averages from all molecules in the asymmetric unit, and a total averaged value is given for the two dArmRPs. The Cα(P/P+2)-distance of 1EE4 is calculated only from one molecule. The total averaged values (rise h, radius r and angle 2·Ω) are schematically represented by a cylinder. The optimal curvature, given by the Cα(P/P+2)-distance of 6.7- 7.0 Å, is found in repeat pair GH of importin-α. The model structure CAR0 was generated by Rosetta based on the curvature of repeat pair GH, and can be described as a thick but short cylinder. Structures of dArmRPs are represented by thin but tall cylinders.

Figure 3. ArmRP sequence alignment of the N-terminal capping repeats (N_V and Y_{III}), internal repeats (D_{SPVA}, Dq, V1, V2, V3, V4 and M) and C-terminal repeats (C_{PAF} and A_{II}). Each designed ArmRP consists of three α-helices (shown as black rectangles), N-terminal caps only have two helices. Residues with alternatives from the computational design approach in protein CAR0 ($N_V(D_{SPVA})_4C_{PAF}$) are highlighted in orange. Mutations introduced based on MD simulations in protein CAR1 ($Nq(Dq)_4C_{PAF}$) are colored grey. Modifications on the binding surface to introduce binding pockets P1' and P2' from importin-α are colored magenta and green (in V1-V4), whereas the charge neutralization mutations are colored in yellow. Sequences from consensus-based ArmRP are highlighted in blue (M refers to the consensus-based internal repeat $\bar{M}$, reported by Alfarano *et al.* [9]).

Figure 4. Biophysical characterization of computationally designed ArmRPs (dArmRPs). A: SEC (normalized absorption at 230 nm) and MALS (dots) of the proteins. Elution volumes of bovine serum albumin (MW: 66 kDa) and carbonic anhydrase (MW: 29 kDa) are indicated by dashed lines and were used as molecular weight standards. B: ANS fluorescence spectra. Introduction of the Y_{III}-cap resulted in ANS signals similar to the reference proteins, shown as horizontal dashed lines, indicating the highest ANS signal observed in the spectra of consensus-based proteins $Y_IM_4A_I$ and $Y_{III}M_4A_{II}$ [10]. C: CD spectra of all proteins are shown, expressed as the mean residue ellipticity (MRE). D: Normalized temperature-induced unfolding of designed proteins (dots) with fits (lines). E: Normalized GdnHCl-induced unfolding of dArmRPs (dots) with fits (lines).

Figure 5. X-ray structures of dArmRP CAR2 ($Y_{III}(Dq)_4C_{PAF}$). A: Both molecules from the asymmetric unit of CAR2 are shown, chain A as a ribbon (colored according to its B-factor (blue, green and red indicate low, medium and high B-factors, respectively) and chain B in surface representation (Y_{III}-, Dq^1 and Dq^3, Dq^2 and Dq^4-, and C-capping-repeats are shown in olive, light grey, dark grey and orange, respectively). The N- and C- termini of each protein are indicated. B: Superposition of the backbone of the CAR2 molecules in the asymmetric unit.

Figure 6. Biophysical characterization of surface-modified dArmRPs binding to peptides. A: SEC (normalized absorption at 280 nm) and MALS (dots) of surface-modified dArmRPs in comparison to

CAR2. B: CD spectra of all proteins, expressed as mean residue ellipticity (MRE). C: Normalized temperature-induced unfolding of designed proteins (dots) with fit (lines). D: Normalized GdnHCl-induced unfolding of dArmRPs (dots) with fit (lines) compared to CAR2 and $Y_{III}M_4A_{II}$. E: Exemplary binding curves of dArmRP and $(KR)_5$-GFP recorded by fluorescence anisotropy (dots) with fits (lines); the corresponding K_d of $Y_{III}M_4A_{II}$ is indicated by a black dotted line. F: Specificity test of dArmRPs by ELISA against different peptides, differing in lengths and charge. The sequences of all peptides are given in Supplementary Table ST4. *No target*, only NeutrAvidin coated. *no protein*, unspecific signals of detection antibodies (in the absence of ArmRP protein).

Figure 7. Peptide binding mode in CAR2.V1. A: Superposition of poly-arginine peptides ($(RR)_5$ shown as blue sticks) on chain A shown in surface representation and colored according to vacuum electrostatics. Binding pockets P1' are occupied by residues Arg2, Arg4, Arg6 and partially Arg8, whereas rationally engineered binding pockets P2' are filled with Arg7, Arg9 and partially Arg5. N- and C-terminal residues of peptides are more flexible and not shown. B: Modular binding mechanism of peptide backbone (salmon stick) of ArmRP chain B. Part of helices H3 are shown as cylinders. Bidentate hydrogen bonds to Asn^{37} and measured Cα(P/P+2)-distances are indicated as yellow and red dashed lines, respectively. Conserved modular backbone binding was observed for residues Arg5 and Arg7. Arg3 and Arg9 are bound in a less conserved manner, indicated by the increased length or complete absence of hydrogen bonds in internal repeat 1 or 4 of chain A-E. C: Conserved binding mode in pocket P2'. Arginine residues are fixed by ionic interactions to Glu^{30} (blue dashed lines) and cation-π interactions with Trp^{30}. Lys^{29} (green) was mutated to Gln in CAR2.V2 and CAR2.V4 to remove the charge neutralization of adjacent Glu^{30}. Plain numbers and superscripts refer to the numbering scheme in the PDB file and individual repeats, respectively. D: Conserved binding mechanism in pocket P1' composed of four hydrogen bonds mediated by Gly^{41*}, Asn^{1} and Ser^{40} (* and $^{\#}$ indicate positions in previous and following repeats, respectively).

Figure 8. X-ray structure of dArmRP CAR2.V1 in complex with $(RR)_5$. A: Tetramer of the asymmetric unit. Two subunits are shown in surface representations (chain A and B) and two as ribbons (chain E and F) and colored dark and light blue, respectively. The N- and C- termini of each protein are indicated (chain names are indicated as subscripts). Each molecule binds a peptide, shown as red sticks, in antiparallel orientation. B: Superposition of the backbone of CAR2.V1 (shown in different

shades of blue) on chain A of CAR2 (grey). C: Superposition of four complexes of CAR2.V1. The dArmRPs are shown as Cα-traces (colored in different shades of blue) and the peptides as sticks in different colors.

Figure 9. Structural details of the computationally designed cap C_{PAF}. A: Superposition of the C_{PAF} cap from the Rosetta model (CAR0, grey) and crystal structure (CAR2, magenta). Major differences are found in the loop region (light magenta). B: Detailed view of the loop conformation mediated by residue His^{22} (yellow). Conserved hydrogen bond to residue Asn^{24} (yellow dashed line) and several hydrophobic interactions (blue dashed lines) are found identically in the internal repeats Dq (shown in grey) by residue Asp^{22} (orange). C: Structural similarity of C_{PAF} superimposed on the C-cap of consensus-based cap A_{II} (green) (PDB ID: 4DB6 [10]).

Figure 10. Structural details of stabilizing mutations introduced by molecular dynamics simulations in internal repeats Dq. A: Mutations introduced in the internal repeats and the C-cap are shown in stick representation in the structure of CAR2. N- and C-cap are colored in orange and green, respectively. B-D: Introduced residues in the Dq sequence are colored in green, and are compared either to the modeled wild-type residue (D_{SPVA}) or another mutation (yellow). Residues in close proximity are shown in grey stick representation (* and # indicate the positions in the preceding and following repeats, respectively). Hydrogen bonds are indicated by yellow dashed lines. Clashes observed in the model wild type structures are shown with red cylinders. B: Gln at position 3 makes three hydrogen bonds in contrast to Met. C: Hydrophobic core mutation V8I fills hydrophobic core completely (Ile show in green). Mutation I8L (a mutation introduced in a CAR2 variant, data not shown) results in several clashes and a reduced protein stability. D: Ile fits into the hydrophobic core at position 4 without clashes.

Fig.1

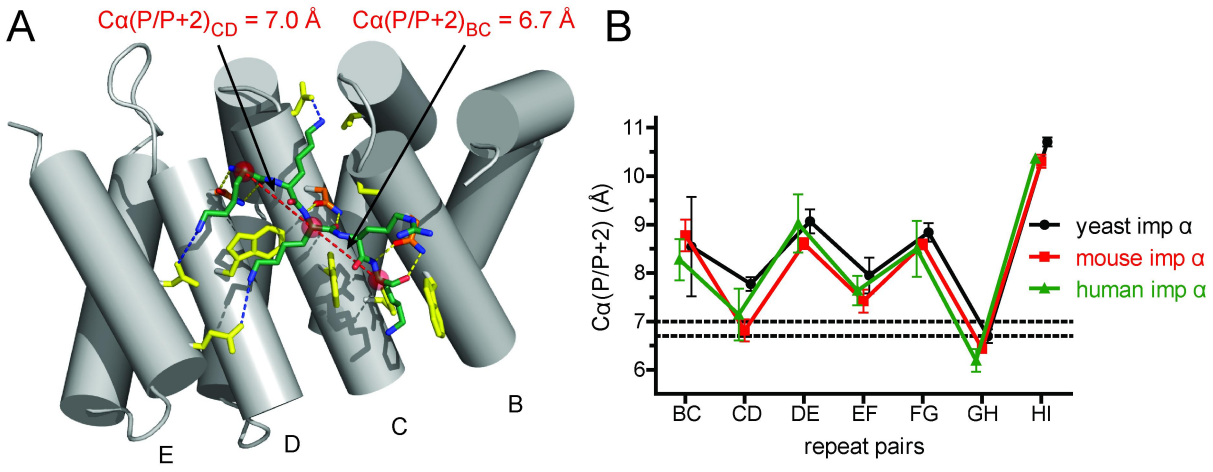
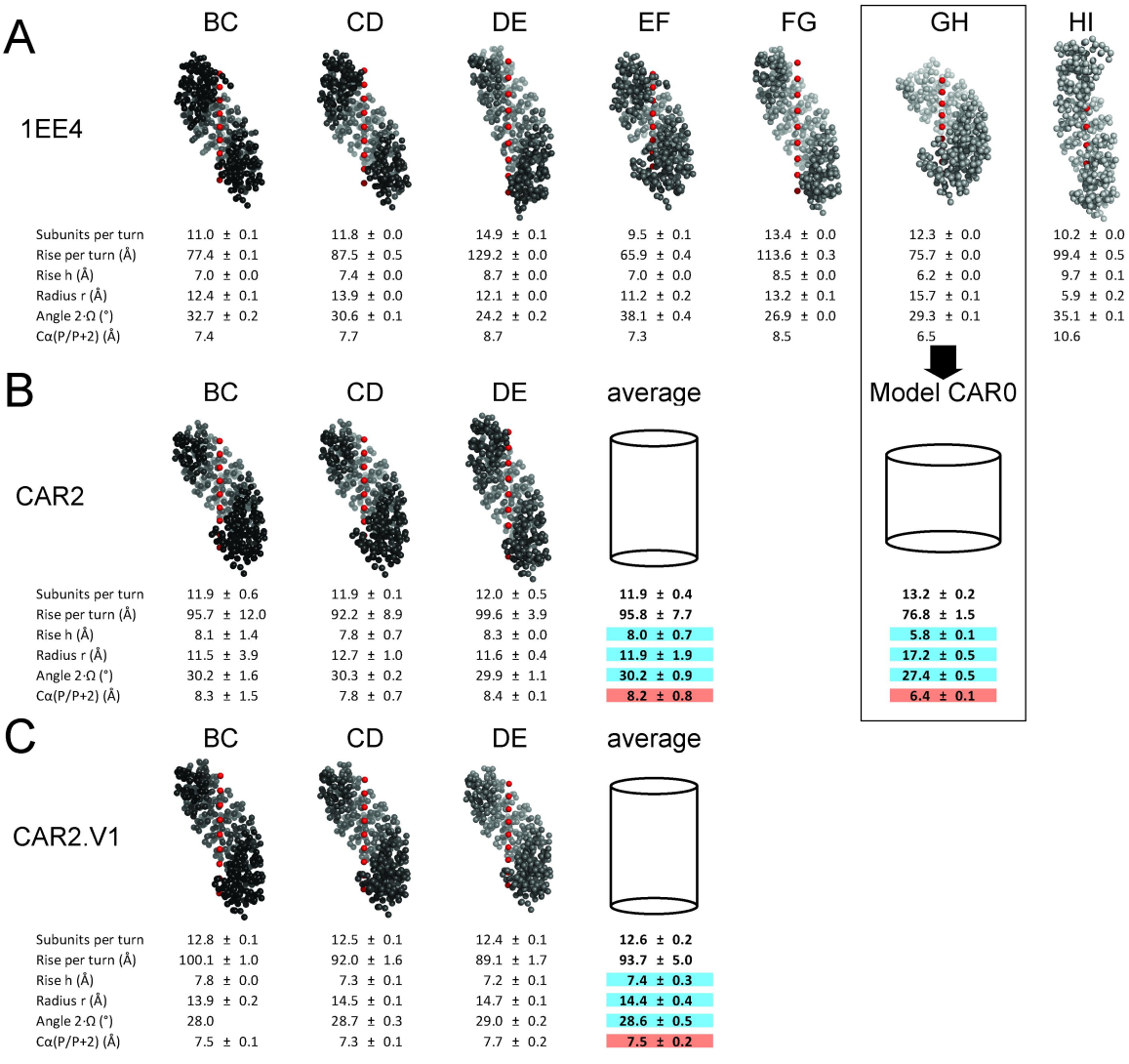


Fig.2



A	BC	CD	DE	EF	FG	GH	HI
Subunits per turn	11.0 ± 0.1	11.8 ± 0.0	14.9 ± 0.1	9.5 ± 0.1	13.4 ± 0.0	12.3 ± 0.0	10.2 ± 0.0
Rise per turn (Å)	77.4 ± 0.1	87.5 ± 0.5	129.2 ± 0.0	65.9 ± 0.4	113.6 ± 0.3	75.7 ± 0.0	99.4 ± 0.5
Rise h (Å)	7.0 ± 0.0	7.4 ± 0.0	8.7 ± 0.0	7.0 ± 0.0	8.5 ± 0.0	6.2 ± 0.0	9.7 ± 0.1
Radius r (Å)	12.4 ± 0.1	13.9 ± 0.0	12.1 ± 0.0	11.2 ± 0.2	13.2 ± 0.1	15.7 ± 0.1	5.9 ± 0.2
Angle 2·Ω (°)	32.7 ± 0.2	30.6 ± 0.1	24.2 ± 0.2	38.1 ± 0.4	26.9 ± 0.0	29.3 ± 0.1	35.1 ± 0.1
Cα(P/P+2) (Å)	7.4	7.7	8.7	7.3	8.5	6.5	10.6

B	BC	CD	DE	average	Model CAR0
Subunits per turn	11.9 ± 0.6	11.9 ± 0.1	12.0 ± 0.5	**11.9 ± 0.4**	**13.2 ± 0.2**
Rise per turn (Å)	95.7 ± 12.0	92.2 ± 8.9	99.6 ± 3.9	**95.8 ± 7.7**	**76.8 ± 1.5**
Rise h (Å)	8.1 ± 1.4	7.8 ± 0.7	8.3 ± 0.0	**8.0 ± 0.7**	**5.8 ± 0.1**
Radius r (Å)	11.5 ± 3.9	12.7 ± 1.0	11.6 ± 0.4	**11.9 ± 1.9**	**17.2 ± 0.5**
Angle 2·Ω (°)	30.2 ± 1.6	30.3 ± 0.2	29.9 ± 1.1	**30.2 ± 0.9**	**27.4 ± 0.5**
Cα(P/P+2) (Å)	8.3 ± 1.5	7.8 ± 0.7	8.4 ± 0.1	**8.2 ± 0.8**	**6.4 ± 0.1**

C	BC	CD	DE	average
Subunits per turn	12.8 ± 0.1	12.5 ± 0.1	12.4 ± 0.1	**12.6 ± 0.2**
Rise per turn (Å)	100.1 ± 1.0	92.0 ± 1.6	89.1 ± 1.7	**93.7 ± 5.0**
Rise h (Å)	7.8 ± 0.0	7.3 ± 0.1	7.2 ± 0.1	**7.4 ± 0.3**
Radius r (Å)	13.9 ± 0.2	14.5 ± 0.1	14.7 ± 0.1	**14.4 ± 0.4**
Angle 2·Ω (°)	28.0	28.7 ± 0.3	29.0 ± 0.2	**28.6 ± 0.5**
Cα(P/P+2) (Å)	7.5 ± 0.1	7.3 ± 0.1	7.7 ± 0.2	**7.5 ± 0.2**

Fig.3

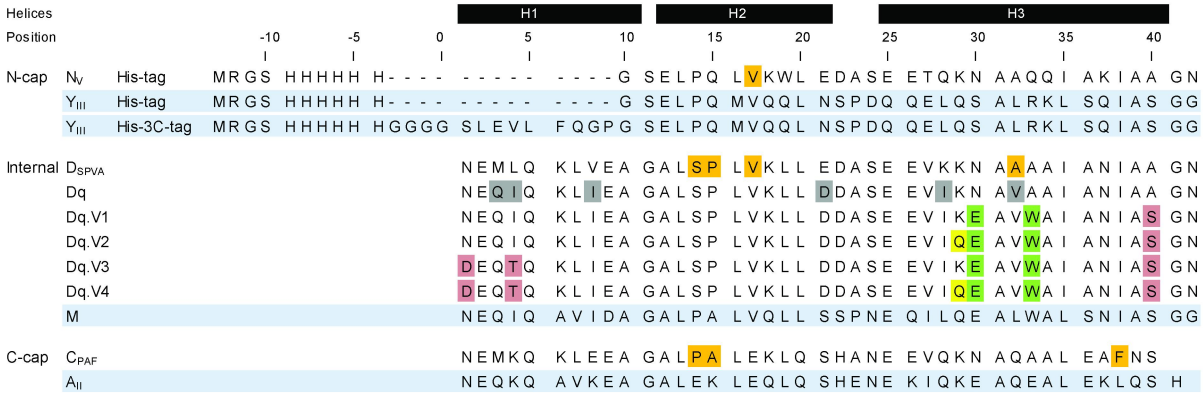

Helices
H1
H2
H3
Position
-10 -5 0 5 10 15 20 25 30 35 40
N-cap
Internal
C-cap
N_V His-tag MRGS HHHHHH H- - - - - - - - - - - - - G SELPQ LVKWL EDASE ETQKN AAQQI AKIAA GN
Y_{III} His-tag MRGS HHHHHH H- - - - - - - - - - - - - G SELPQ MVQQL NSPDQ QELQS ALRKL SQIAS GG
Y_{III} His-3C-tag MRGS HHHHHH HGGGG SLEVL FQGPG SELPQ MVQQL NSPDQ QELQS ALRKL SQIAS GG
D_{SPVA} NEMLQ KLVEA GALSP LVKLL EDASE EVKKN AAAAI ANIAA GN
Dq NEQIQ KLIEA GALSP LVKLL DDASE EVIKN AVAAI ANIAA GN
Dq.V1 NEQIQ KLIEA GALSP LVKLL DDASE EVIKE AVWAI ANIAS GN
Dq.V2 NEQIQ KLIEA GALSP LVKLL DDASE EVIQE AVWAI ANIAS GN
Dq.V3 DEQTQ KLIEA GALSP LVKLL DDASE EVIKE AVWAI ANIAS GN
Dq.V4 DEQTQ KLIEA GALSP LVKLL DDASE EVIQE AVWAI ANIAS GN
M NEQIQ AVIDA GALPA LVQLL SSPNE QILQE ALWAL SNIAS GG
C_{PAF} NEMKQ KLEEA GALPA LEKLQ SHANE EVQKN AQAAL EAFNS
A_{II} NEQKQ AVKEA GALEK LEQLQ SHENE KIQKE AQEAL EKLQS H

Fig.4

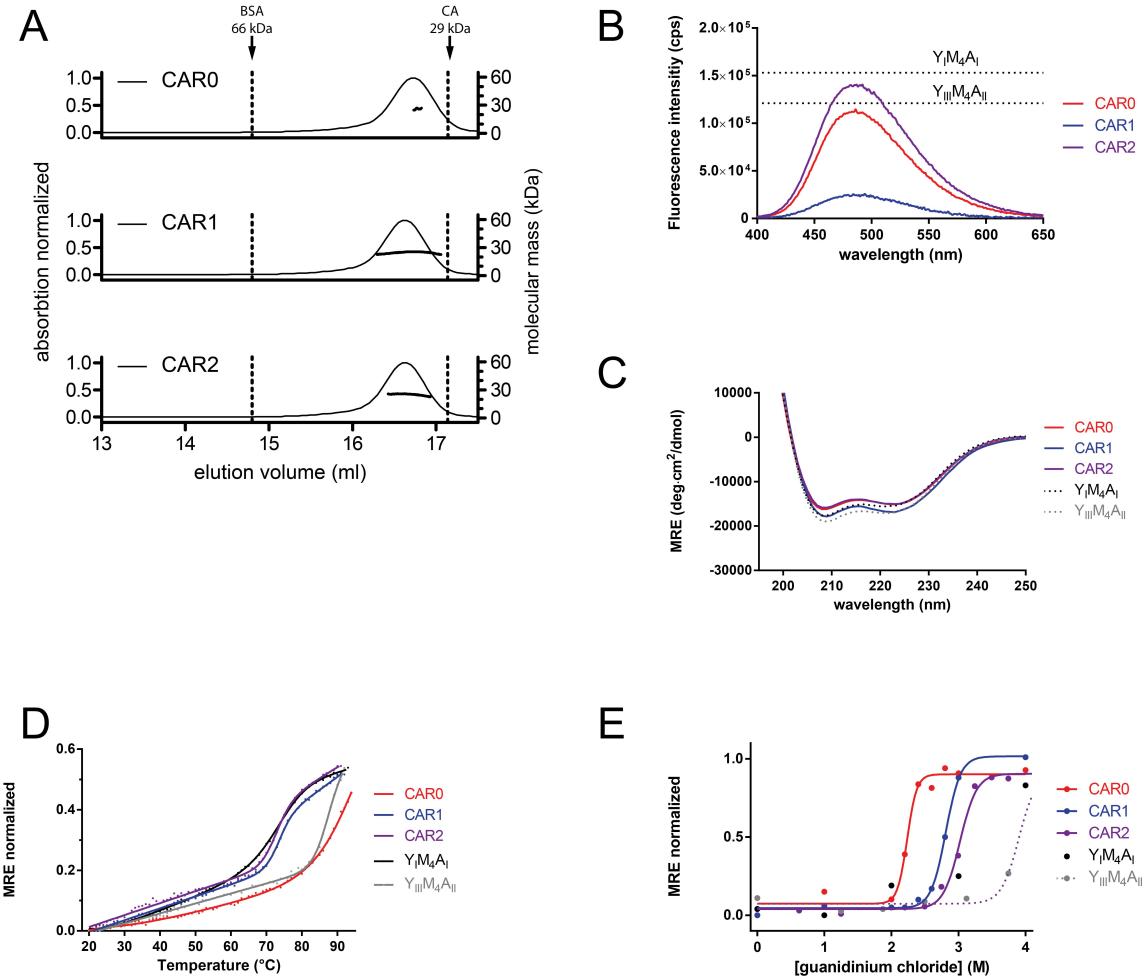

A
BSA 66 kDa
CA 29 kDa
CAR0
CAR1
CAR2
absorbtion normalized
molecular mass (kDa)
elution volume (ml)
B
$Y_IM_4A_I$
$Y_{III}M_4A_{II}$
Fluorescence intensity (cps)
wavelength (nm)
C
MRE (deg.cm²/dmol)
wavelength (nm)
D
MRE normalized
Temperature (°C)
E
MRE normalized
[guanidinium chloride] (M)

Fig.5

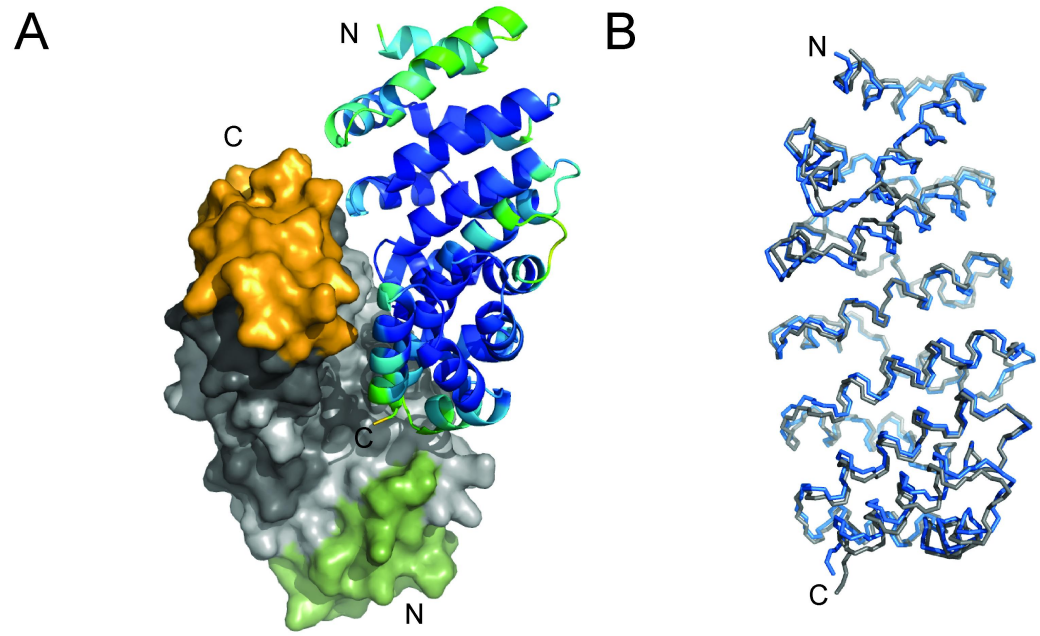

A
N
C
C
N
B
N
C

Fig.6

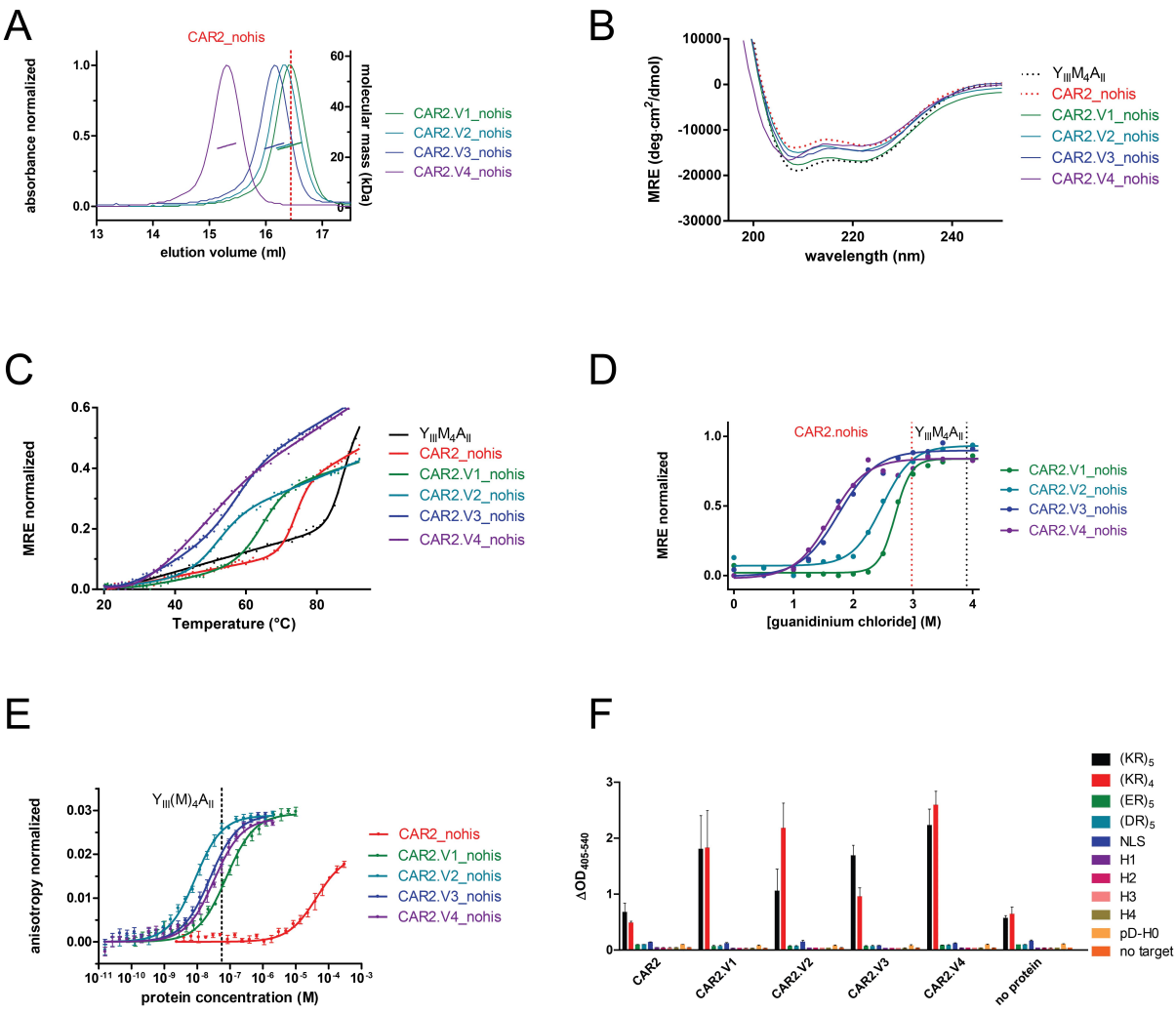
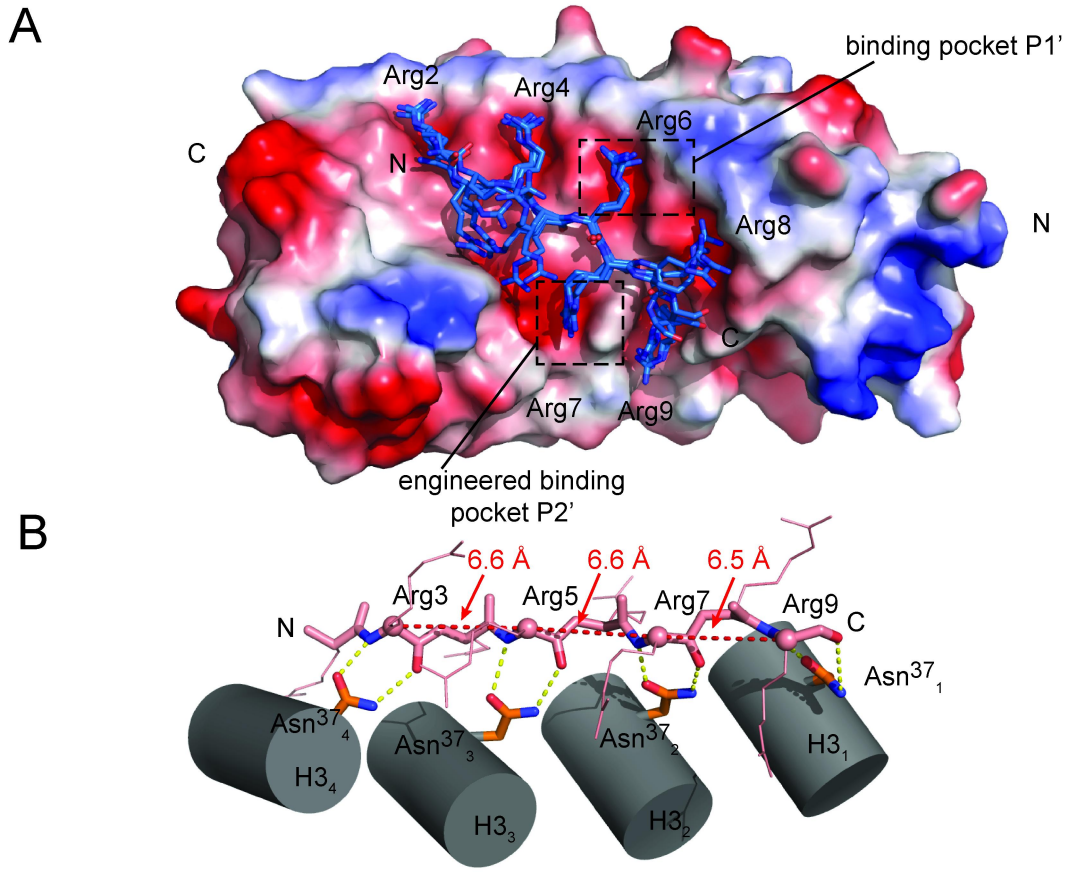

A
CAR2_nohis
absorbance normalized
molecular mass (kDa)
elution volume (ml)
CAR2.V1_nohis
CAR2.V2_nohis
CAR2.V3_nohis
CAR2.V4_nohis
B
MRE (deg·cm²/dmol)
wavelength (nm)
YIIIM4AII
CAR2_nohis
CAR2.V1_nohis
CAR2.V2_nohis
CAR2.V3_nohis
CAR2.V4_nohis
C
MRE normalized
Temperature (°C)
YIIIM4AII
CAR2_nohis
CAR2.V1_nohis
CAR2.V2_nohis
CAR2.V3_nohis
CAR2.V4_nohis
D
CAR2.nohis
YIIIM4AII
MRE normalized
[guanidinium chloride] (M)
CAR2.V1_nohis
CAR2.V2_nohis
CAR2.V3_nohis
CAR2.V4_nohis
E
YIII(M)4AII
anisotropy normalized
protein concentration (M)
CAR2_nohis
CAR2.V1_nohis
CAR2.V2_nohis
CAR2.V3_nohis
CAR2.V4_nohis
F
ΔOD405-540
CAR2
CAR2.V1
CAR2.V2
CAR2.V3
CAR2.V4
no protein
(KR)5
(KR)4
(ER)5
(DR)5
NLS
H1
H2
H3
H4
pD-H0
no target

Fig.7



	Arg3		Arg5		Arg7		Arg9	
	HB1*	HB2**	HB1*	HB2**	HB1*	HB2**	HB1*	HB2**
chain A	3.1	2.8	2.8	3.0	3.4	3.3	n.a	n.a
chain B	2.9	2.4	3.0	3.0	2.8	2.6	3.5	3.0
chain E	3.1	3.0	3.1	3.1	3.0	2.8	n.a	n.a
chain F	3.5	3.8	3.0	2.9	2.6	2.8	3.1	2.7

* HB1: hydrogen bond distance (Å) between Asn^{37}_{X} ND2 of ArmRP and ArgY- O of peptide (X: number of IR, Y: bound peptide residue)

** HB2: hydrogen bond distance (Å) between Asn^{37}_{X} OD1 of ArmRP and ArgY- N of peptide (X: number of IR, Y: bound peptide residue)

Color code: dark green: strong HB, medium green: medium strong HB, light green, weak HB.

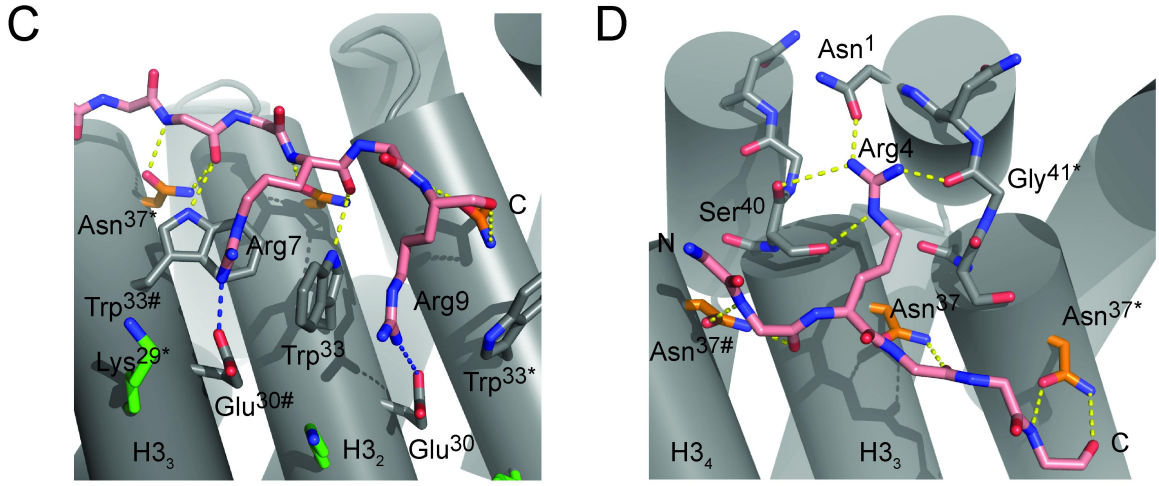


Fig.8

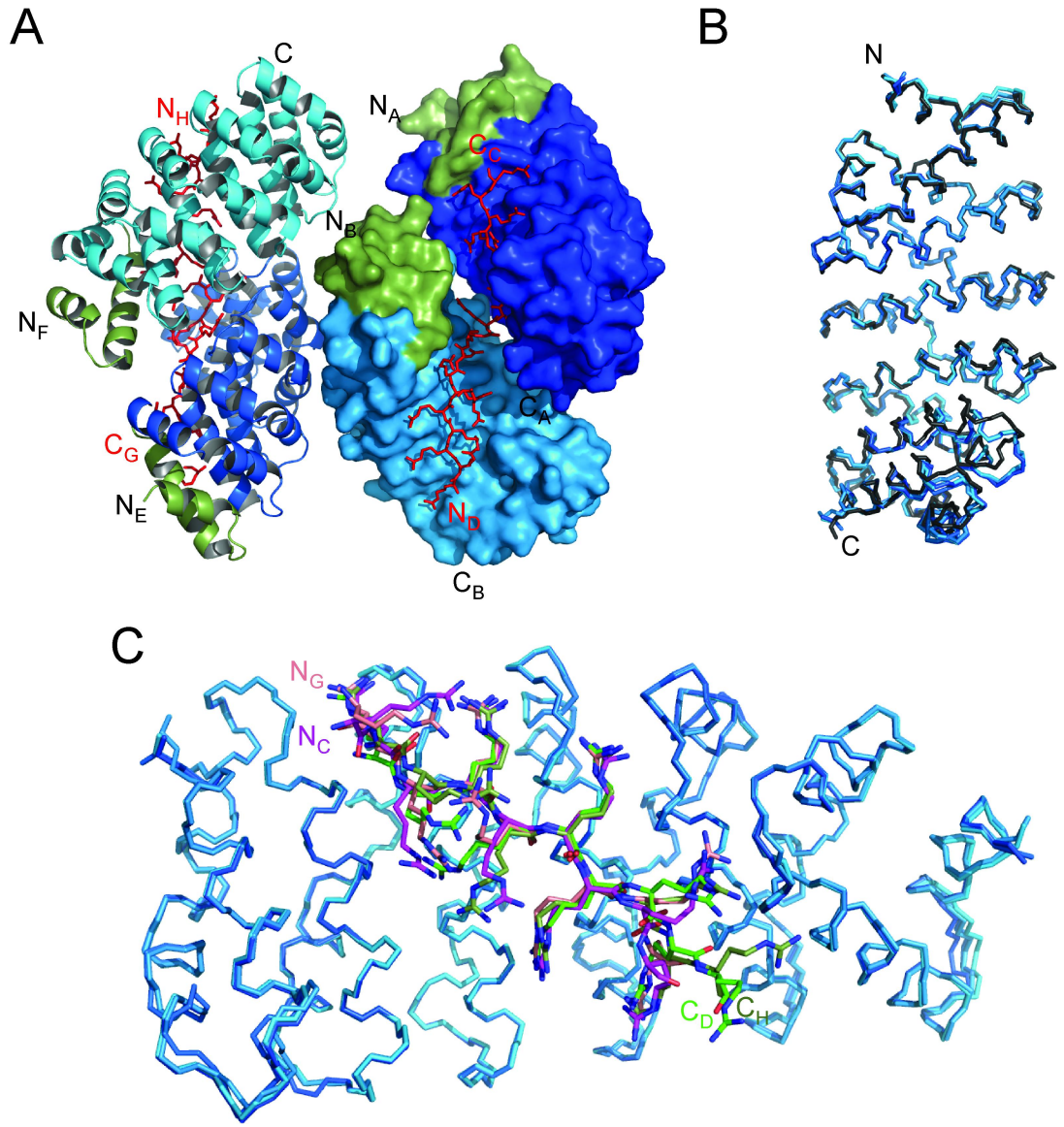

A
C
N_H
N_A
C_C
N_B
N_F
C_A
C_G
N_E
N_D
C_B
B
N
C
C
N_G
N_C
C_D
C_H

Fig.9

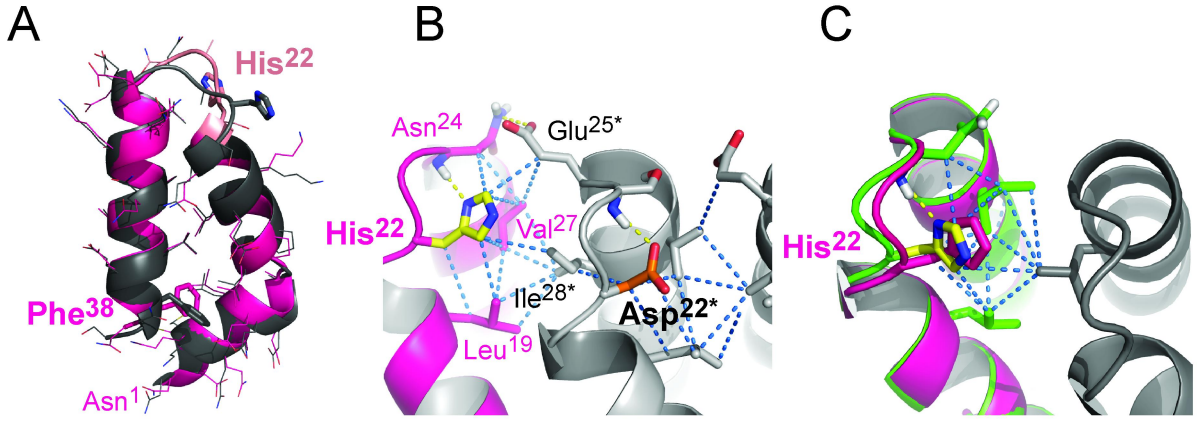
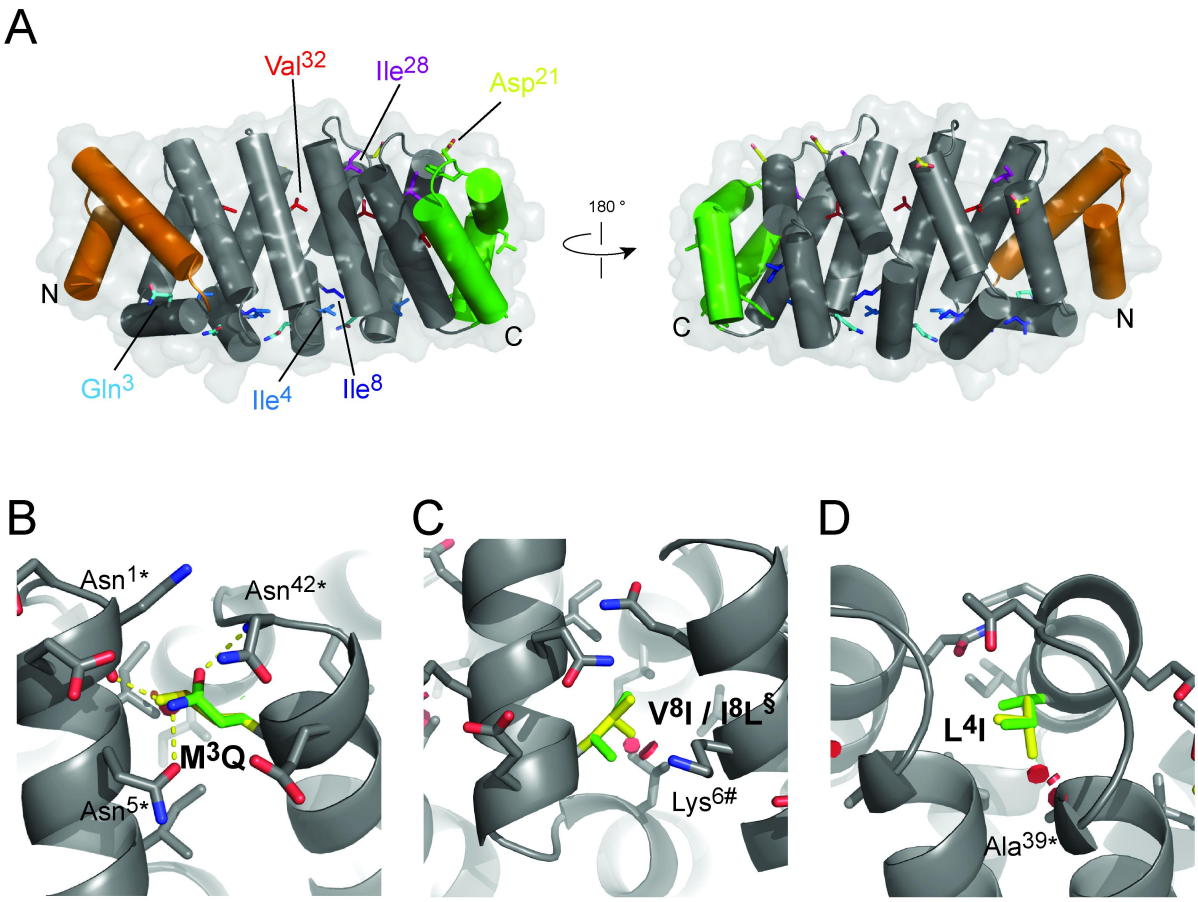

A
His22
Phe38
Asn1
B
Asn24
Glu25*
His22
Val27
Ile28*
Asp22*
Leu19
C
His22

Fig.10


A
Val32
Ile28
Asp21
180 °
N
C
C
N
Gln3
Ile4
Ile8
B
Asn1*
Asn42*
M3Q
Asn5*
C
V8I / I8L§
Lys6#
D
L4I
Ala39*